To: Unicode Technical Committee
From: Debbie Anderson, Script Encoding Initiative, UC Berkeley
Subject: Summary of email discussion on Gurmukhi BINDI
Date: 17 February 2020

There was a Gurmukhi Ad Hoc held during the Unicode Technical Committee meeting in October 2019. Various options were explored at the meeting. The group agreed that if the left-side BINDI is required for plain text, then additional evidence is required.

Below is a summary of an email thread on Gurmukhi BINDI, with comments from Kulpreet Chilana, Irvanjit Singh, and Manvir Singh, dating from October 2019 - January 2020. In the comments from Manvir Singh and Irvanjit Singh, Kulpreet Chilana's responses are italicized and indented.

Kulpreet Chilana:
Based on the chapter of Gurbani linguistics and grammar by Mangat Rai Bhardwaj in his *Colloquial Punjabi II* (see ADDENDUM 1, below), Gurmukhi was originally written in *larivaar* (i.e., text written in continuous lines without spaces between words), and *bindi* was arbitrarily placed in many manuscripts to compensate for clashing that would otherwise occur.
This leads me to believe:
(1) the position of the *bindi* is not orthographically or linguistically significant—even in the context of Gurbani;
(2) avoidance of this clashing can happen at that font-level when Gurbani is rendered in the *larivaar* form.

If evidence is provided with *bindi* and *tippe* being used consistently in same position, the following are options:
(1) Encode two new code points for *LEFT SIDE BINDI* and *LEFT SIDE TIPPE.*

Pitfalls:
(a) will require support be added to Gurmukhi fonts, which will take time to be picked up by vendors;
(b) would require all existing Gurmukhi databases be updated;
(c) the Gurmukhi block will be polluted with multiples of the same character rendered in different contexts;
(d) it will be confusable to users, i.e., which to use of the following:

ਕ + ਿ + ਤ + BINDI + ੀ or ਕ + ਿ + ਤ + ੀ + LEFT SIDE BINDI

(2) Use a variation selector with *bindi* and *tippe* to indicate it should be rendered on the left side. Similar pitfalls as those above, but wouldn't create confusion for users if both versions of *bindi* and *tippe* appeared on keyboard. The Script Ad Hoc did not seem keen in adopting this.

[(3) I am supportive of your idea to break down Gurmukhi vowels into components, (i.e. ੲ + ੀ = ਈ), assigning this sort of equivalence is evidently impossible because of Unicode stability and other security considerations.]

Manvir Singh:
Mangat Rai Bhardwaj's *Colloquial Punjabi II* only briefly mentions *bindi*. His comment on the

arbitrary positioning of *bindi*s by scripts doesn't make much sense for a script like Gurmukhi. In Arabic, nuqtey often shift position to look nicer or avoid too much clutter in an area. But such an issue doesn't occur in Gurmukhi, regardless of whether they are in *larivaar* or *padchhed* (text in which words are separated with spaces), …however [in] a lot of cases … *Bindi* [that] occur before Bihari occur independently of where the text is located on a page. This is especially evident in the Shabad where almost every line ends with ਏਂੀ.

> *Kulpreet: Unlike other Indic scripts, Gurmukhi has a strong calligraphy tradition that is heavily influenced by Persian nastaliq. In general, I think it's plausible that scribes adjusted the position of the bindi similar to how they would if they were writing in Persian—but there's no way to say for sure.*

I attach evidence of *bindi* in the same position consistently (see ADDENDUM 2). I have just stuck to looking for the same Shabad. You'll see that though a few are missing *bindi,* the ones that have the *bindi* before the Bihari for this Shabad in particular.

#6 has *tippis* instead (which I believe I have addressed in an earlier message about *bindis* and *tippis* being used interchangeably in some texts) and even the *tippi* appears before the Bihari for this Shabad.

#8 shows one of the *bindis* on top of the *kanna*, whereas the second one is near the Bihari. This may be related to the note about *bindi* in this position applying to multiple vowels when between them like that, but I'm not too sure if that's the case.

#22 is very notable, as it is from a manuscript by Bhai Mani Singh, the scribe of the original Damdami Beerh of SGGS. In this example, he not only substitutes in a *tippi* for *bindi*, but the *tippi* appears inside the Bihari. This seems to imply that there is some unique reason as to why the *tippi* is appearing before the Bihari as opposed to after

> *Kulpreet: No one can pinpoint why bindi appears before Bihari. As shown in the examples, its usage is inconsistent across different manuscripts – sometimes appearing as a tippe, sometimes appearing to the lef of the Bihari, sometimes not appearing at all. #22 makes me more confident that this should be accomplished at the font-level; perhaps a ligature when you type Bihari + tippe, that draws a tippe in the Bihari as it appears in the original SGGS or renders it as the left-side bindi.*

Irvanjit Singh:
The primary purpose of the Gurmukhi script when it was created was to provide a simplified means of preserving Gurbani in written form. Thus, any discussion on the Gurmukhi script should be framed in the context of this purpose. To this end, the preservation of Gurbani *as it is written* is of paramount importance, from both scholarly and spirituality-based perspectives.

This is the approach that scribes have taken when writing Beerhs of SGGS and evidently, it is the approach that Dr. Kulbir Singh Thind took when he created the AnmolLippi and GurbaniAkhar typefaces and Gurbani CD. This has also been done within reason, and almost all of the glyphs in Dr. Thind's font are already encoded in the Gurmukhi block. That being said, if there is sufficient evidence for a missing character such as the Left Side Bindi, our approach will always be to render it as such in our fonts, regardless of whether we fully understand its significance at the time.

Kulpreet: As you mentioned, the primary purpose of the Gurmukhi script was to create a simplified means of preserving spoken Gurbani. This is true—however, a simplified script that's more accessible comes with the tradeoff of making the script less expressive. …As you know, Gurbani is comprised of many languages and creating a script that accurately represents all of the phonemes in these languages would make it very complex…At best, the Gurmukhi script is used as an approximation for how it may have been pronounced, but there's really no way of knowing how it was pronounced when it was written… The ambiguity that's inherent in the Gurmukhi script leads me to believe that that the inventors (the first and second Sikh gurus) intentionally did not place emphasis on the script as the ultimate means of preserving the pronunciation of Gurbani.

It is quite likely that Gurbani (mostly from SGGS) is the most frequently quoted Gurmukhi text on social media. Many Sikh organizations regularly add Gurbani passages to their social media posts. I believe that this has become one of the most common mediums that many youth now engage with Gurbani. The issue with this is that the Gurmukhi unicode block is effectively incomplete without a solution for the Left Side Bindi (among a few other issues), and without one of the standardized solutions mentioned by Kulpreet Singh, there is no way to guarantee that Gurbani will be rendered on users screens as intended. This is a serious concern for Sikhs.

[Discussion on this thread has moved on to] justifying the *reasoning for* the character's use. It seems that regardless of some inconsistencies between a *Tippi* and *Bindi* (bearing in mind potential interchangeability) or the specific position (inside the bihari or outside it), there is consistency in the character's use as a *left-side variant of the original character*. Scholars may very well never know with complete certainty what the actual reasoning for this may have been, but as mentioned [above], even if in the future it is determined that the position of the *bindi* has no orthographical or linguistic significance, this will not change our approach to preserving the variations in our fonts.

Kulpreet: Establishing linguistic or orthographic significance is important to determining where the left side bindi fits in Gurmukhi block and what technical solution is appropriate and correct. In this case, we're trying to determine if the left-side bindi is in-fact distinct from TIPPE or BINDI—and no evidence seems to support that it is. That being said, I understand the need to arrive at a standardized solution for this problem so people can read and share Gurbani and expect it to be rendered the same across platforms.
Based on the evidence, I make the following proposal:
We've seen evidence of the TIPPE being used inside the BIHARI to represent what later became the left-side bindi. Of particular note is Bhai Mani Singh's original manuscript. We've also seen evidence of the interchangeability to the TIPPE and BINDI. We also know that a TIPPE followed by ਉ ਊ ੀ ੇ ੈ ੋ ੌ (and dependent variants) is orthographically incorrect in Gurmukhi. Thus, I propose that we recommend ਉ ਊ ੀ ੇ ੈ ੋ ੌ (and dependent variants) + TIPPE be rendered in the font with a left-side bindi. This recommendation should be included in the next version of the Gurmukhi chapter in the Unicode standard.

Manvir Singh: I am not sure that your solution of using a Tippi to render a left-side Bindi is the best solution. The solution seems a bit hacky to me, so to speak. It doesn't make sense for a Tippi to be used for left side Bindi.

> *Irvanjit Singh: I agree with what Manvir Singh has said [above], the solution you're proposing doesn't seem to make much sense. It isn't intuitive and could also be confusing.*

The issue of cluttering the Gurmukhi unicode block is valid, however, there have been some inconsistencies to this consideration in the past: Namely with the addition of U+0A76. This character could have been implemented at the font level, but it was assigned a code point all the same. There is also the question of the code points used to represent ਈ ਐ ਉ, etc. (and arguably ਸ਼, ਲ਼, and the other ਪੈਰੀਂ-ਬਿੰਦੀ variants of the standard letters). The Gurmukhi block is already cluttered. Yes, this shouldn't get worse, but in this case I think the additions would be justified. I think both code points and selector variables are reasonable solutions.

Regarding the concerns of multiple orderings being confusing for the code point solution, it seems to me that BINDI + BIHARI should remain invalid as per unicode rules, and the only possible combination would be BIHARI + LEFT SIDE BINDI, but perhaps I'm missing some nuances here.

> *Kulpreet: Regardless of what's considered valid Unicode—we have to be careful that users don't input these invalid sequences. I can imagine a situation where regulation in India makes it compulsory to include all characters in the Gurmukhi block on software keyboards (this has happened in the past) and now users are confused as to which character to tap on the screen.*

[Later comments from Irvanjit Singh]
The point [Kulpreet] raised about potential confusion for Indian users if both versions were to be on a keyboard can be mitigated by only allowing LEFT SIDE BINDI to be placed before a BIHAARI and not after. If that doesn't make sense then LEFT SIDE BINDI can be made to only render if placed after a BIHAARI as that is the only current use case. This would help with avoiding confusion. If that doesn't make sense then perhaps variation selectors for BINDI and TIPPI would be the best option. I'm curious to know why the committee isn't in favour of this solution. We already use variation selectors extensively for other characters.

I want to clarify that of the three solutions that Kulpreet Singh has suggested thus far, both Manvir Singh and I find the first two to be perfectly acceptable (hopefully they can be tweaked to be more ideal). To summarize, these were:

1. Encode LEFT SIDE BINDI and LEFT SIDE TIPPI as separate code points (potentially with varying character precedence and rendering exceptions?)

2. Use Variation selectors with BINDI and TIPPI to render left-side variants of these characters.
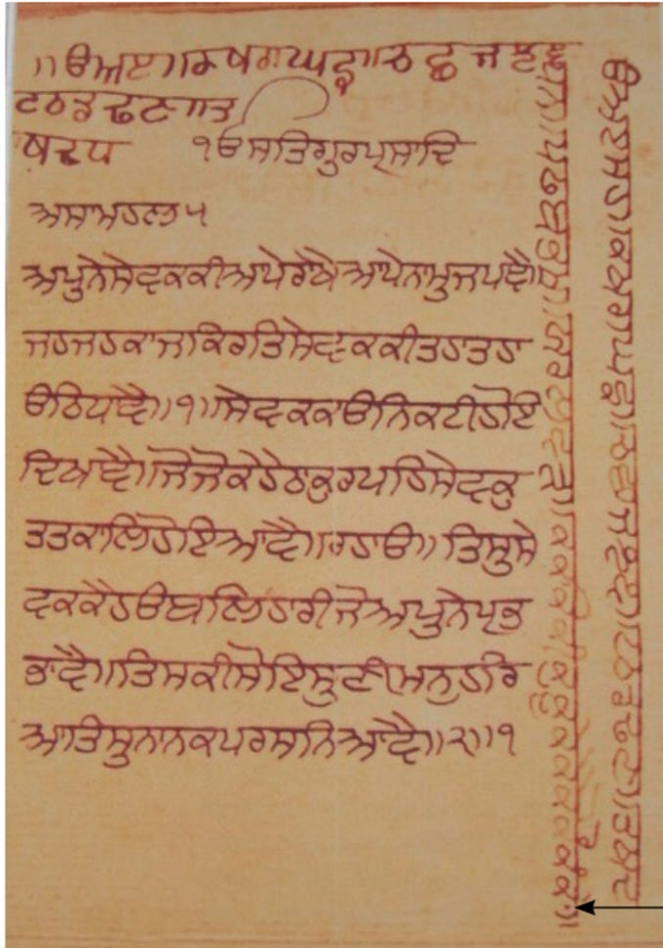
The position of the Bindi is notable

6. The position of the Bindi

As there was originally no ਪਦਛੇਦ in the writing of Gurbani, the scribes put the Bindi where they found it convenient. So we find inconsistency in placing this symbol, either before or after the vowel symbol. Examples are:
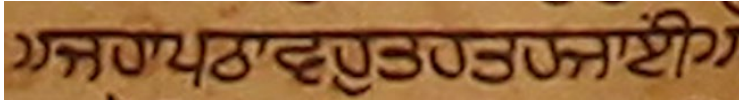
ਈਂਟ, ਧਿਆਈਂ, ਕਿਤੀਂ, ਉਂਥੈ, ਸਕਂਉ

But this seems to be linguistically insignificant. Look at the position of the Bindi at the end of the second downward coming line in the picture of Guru Arjan Dev's handwriting on page 155.

5

ADDENDUM 2: Evidence provided by Manvir Singh

Text:

ਜਹਾ ਪਠਾਵਹੁ ਤਹ ਤਹ ਜਾਈਂ ॥

#1

#2 (?)
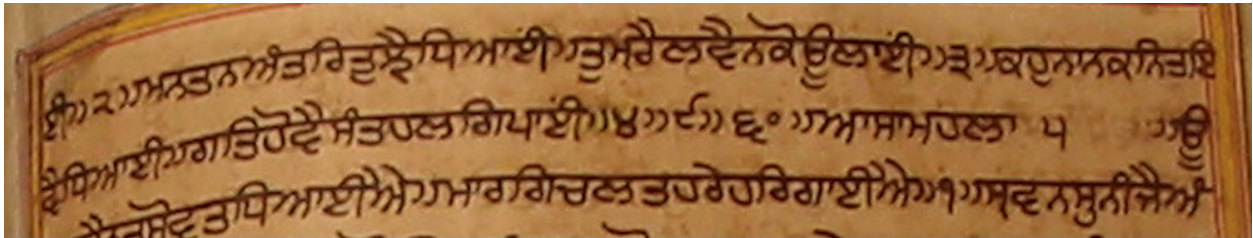
#3
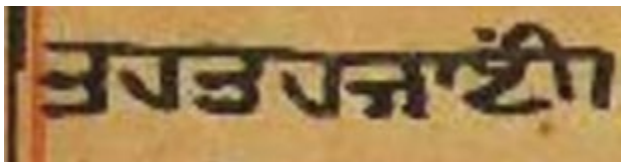
#4 missing Bihari
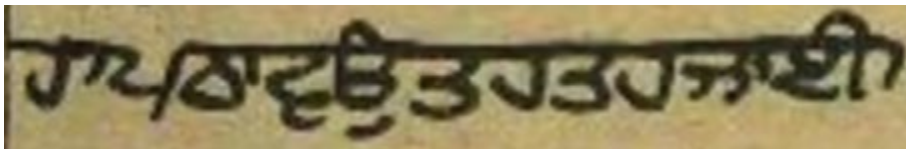
#5
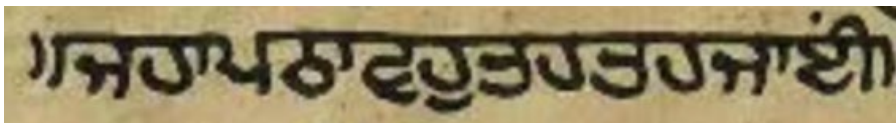
#6 with TIPPI instead of BIHARI:
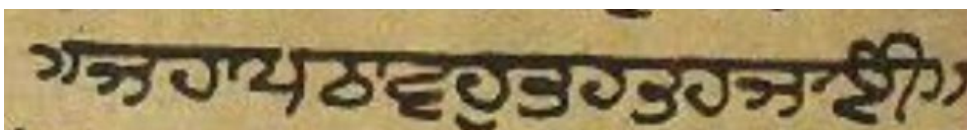
#7 (appears to be missing BINDI)
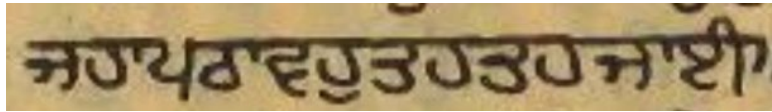
#8 with BINDI on top of KANNA (○ᵀ 0A3E VOWEL SIGN AA)
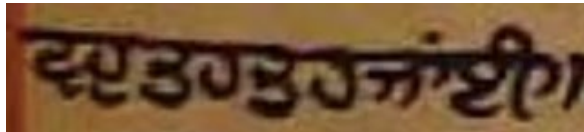
Cf. position of BINDI elsewhere in the same document:

#9

#10

#11

#12

#13

#14

ਤਧਤੁਹਜਾਈਂ

#15

ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ

#16

ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ

#17

ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ

#18

ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ

#19

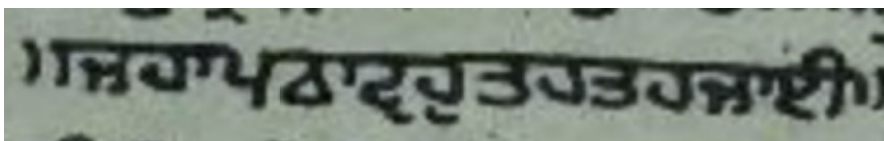ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ
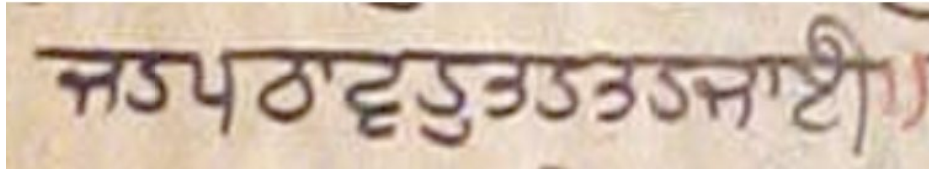
#20

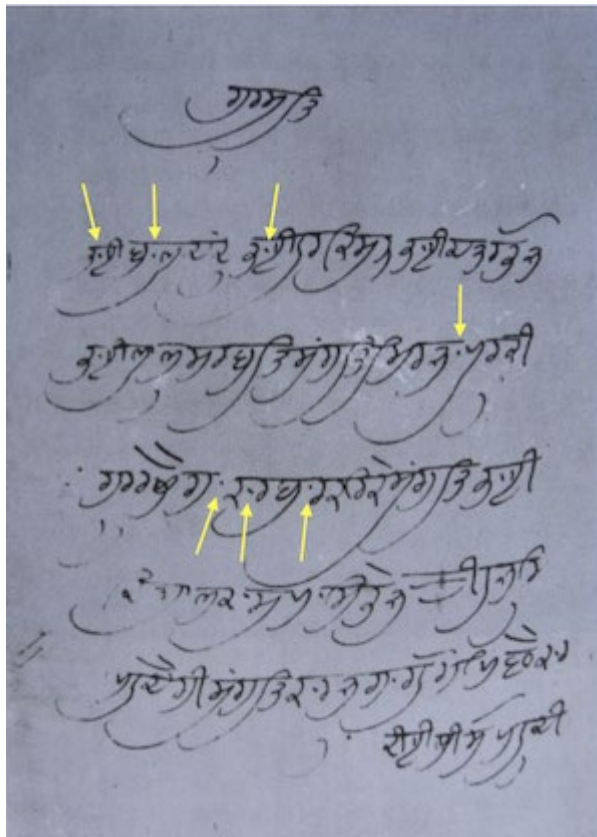ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ

#21

ਯਜਹਾਪਠਾਵਹੁਤਹਤਹਜਾਈਂ
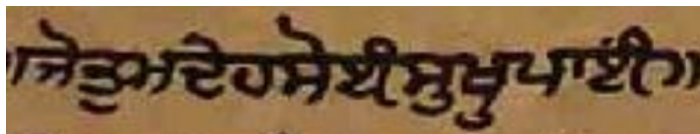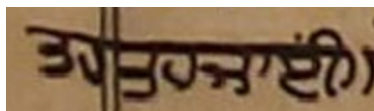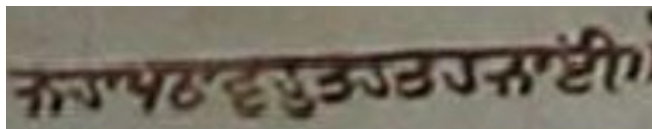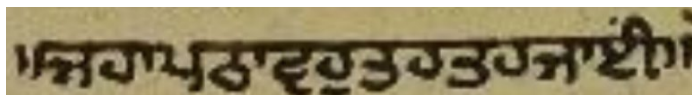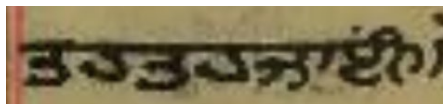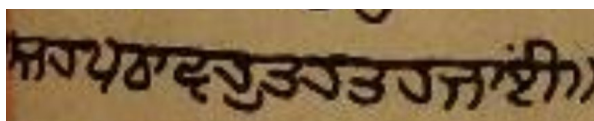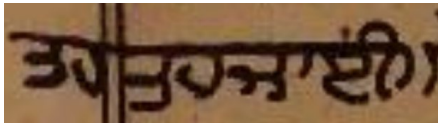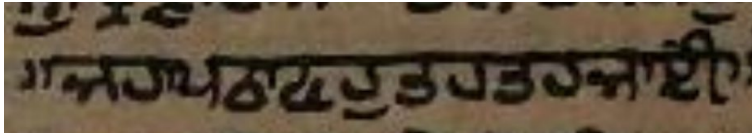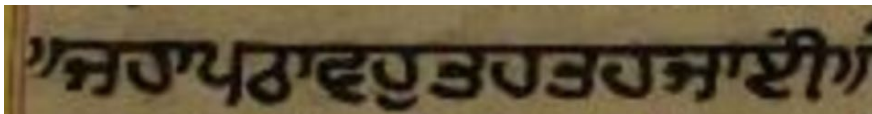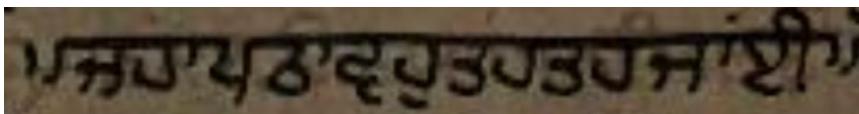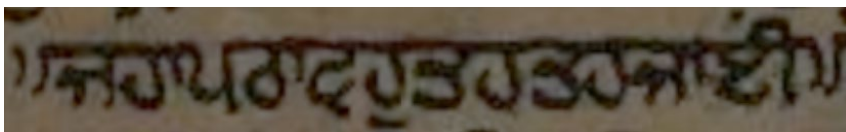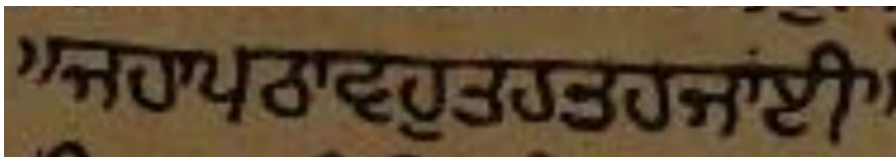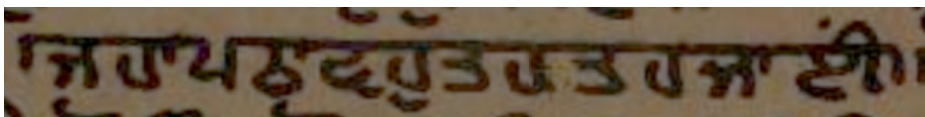
#22 (shows TIPPI for BINDI located inside BIHARI)

ADDENDUM 3: KANNA comment from Manvir Singh

In some old handwritten texts, a dot in the position shown in the image often stands for a *kanna*
(○T 0A3E VOWEL SIGN AA). See the following.

Even Guru Gobind Singh's own handwritten Bani has *Kanna* shown like that. Anyways, that is a bit of a different topic, but I thought it might have been worth mentioning. The example shown in the book is a bit confusing since a *bindi* isn't usually used on a character alone and usually is attached to a vowel sign, so I'm not too sure if it is meant to be a regular *bindi* or not. Of course it is plausible considering that the line appears to show all the vowel signs (though *kanna* kind of classes with the line of Sihari).
(Note from Kulpreet: This is a valid consideration. In this particular manuscript the kanna is not written as a dot, which is evident from the text on the left.)

Regarding the alternate appearance of *kanna*, I've recently come across the fact that many manuscripts use that form of *kanna* when there is not room left on the line for a full *kanna*. I discussed this with an expert on Sikh literature and he confirmed that this is standard for manuscripts of Sikh texts such as Sri Guru Granth Sahib. I guess that explains the *Bindi* on the side in that picture provided by Mangat Rai Bhardwaj.