Distinct orthography boosts morphophonological discrimination: Vowel raising in Bengali verb inflections

Nadja Althaus [a,b,*,1], Sandra Kotzor [b,c,1], Swetlana Schuster [b,2], Aditi Lahiri [b]

[a] *University of East Anglia, Norwich, United Kingdom*
[b] *University of Oxford, Oxford, United Kingdom*
[c] *Oxford Brookes University, Oxford, United Kingdom*

ABSTRACT

This study is concerned with how vowel alternation, in combination with and without orthographic reflection of the vowel change, affects lexical access and the discrimination of morphologically related forms. Bengali inflected verb forms provide an ideal test case, since present tense verb forms undergo phonologically conditioned, predictable vowel raising. The mid-to-high alternations, but not the low-to-mid ones, are represented in the orthography. This results in three different cases: items with no change (*NoDiff*), items with a phonological change not represented in the orthography (*PronDiff*) and items for which both phonology and orthography change (*OrthPronDiff*).

To determine whether these three cases differ in terms of lexical access and discrimination, we conducted two experiments. Experiment 1 was a cross-modal lexical decision task with auditory primes (1st PERSON and 3rd PERSON forms, e.g. [lekʰe] or [likʰi]) and visual targets (verbal noun; e.g. [lekʰa]). Experiment 2 uses eye tracking in a fragment completion task, in which auditory fragments (first syllable of 1st or 3rd PERSON form, e.g. [le-] from [lekʰe]) were to be matched to one of two visual targets (full 1st and 3rd PERSON forms, [lekʰe] vs. [likʰi] in Bengali script).

While the lexical decision task, a global measure of lexical access, did not show a difference between the cases, the eye-tracking experiment revealed effects of both phonology and orthography. Discrimination accuracy in the *OrthPronDiff* condition (vowel alternation represented in the orthography) was high. In the *PronDiff* condition, where phonologically differing forms are represented by the same graphemes, manual responses were at chance, although eye movements revealed that match and non-match were discriminated. Thus, our results indicate that phonological alternations which are not represented in spelling are difficult to process, whereas having ortho-graphically distinct forms boosts discrimination performance, implying orthographically influenced mental phonological representations.

1. Introduction

In speech processing, listeners match the auditory information in the signal to a stored representation (phonological form) in their lexicon. However, in many cases, the phonological information extracted from the signal is not an exact match for the phonological form of the stored representation. Nevertheless, alternating vowels in pairs such as *mean* [miːn]/*meant* [mɛnt] and *sane* [seɪn]/*sanity* [sænɪti] activate the same lexical entry of their underlying base, despite the differences in vowel (Kiparsky, 1982). In English, these alternations are restricted to certain suffixes. However, vowel alternations are a wide-spread phenomenon occurring in verb paradigms across languages, not just within irregular but also within regular verbs (e.g. German strong verbs, e.g. *treffe* [tʁɛfə] 'meet-1P.SG' ~ triffst [tʁɪfst] 'meet-2P.SG') as well as other syntactic cat-egories (e.g. English derivation, *serene* ~ *serenity*; see Fowler, Napps, & Feldman, 1985; Marslen-Wilson, Tyler, Waksler, & Older, 1994). In

addition to the phonological processes involved in recognising such a form (i.e. activating the relevant representation in the lexicon, but also discriminating it from grammatically similar forms), orthography adds another dimension. Since inflected forms are often the outcome of a diachronic process where spelling reflects pronunciation in an earlier form of the language, orthography can be unsystematic. For instance, in Dutch the vowel length difference between *dag* [dax] 'day-sɢ' ~ *dagen* [da:ɣen] 'day-ᴘʟ' is not reflected in the writing, nor is the difference between English *read* [riːd] 'read-ᴘʀᴇs' and *read* [rɛd] 'read-ᴘᴀsᴛ', whereas the phonologically similar difference *lead* [liːd] 'lead-ᴘʀᴇs' ~ *led* [lɛd] 'lead-ᴘᴀsᴛ' does involve a change in spelling. Do vowel changes, as well as the orthographic complexities, make resolving the form more difficult than if the vowel remains the same as in the base form? If there is no cost resulting from the discrepancy between surface form and lexical representation, such alternations could offer a processing advantage as the disambiguation of distinct forms could begin earlier – as soon as the stem vowel is encountered, rather than on hearing the suffix (cf. Lahiri & Marslen-Wilson, 1991). The roles of phonology and orthography in resolving spoken morphological forms has not yet been investigated systematically. One possibility is that orthographic representations are activated immediately when processing speech, and therefore differential spelling may imply more distinct mental activation patterns (e.g. Frost & Ziegler, 2007). Here we assume phonological information is used as soon as it is decoded, i.e. the lexical representation will be activated as soon as possible. The lexical representation consists of the phonological and orthographic forms of the stem and information of all possible affixes and their phonological exponents as well as semantic and syntactic information (see Plank & Lahiri, 2015, for detail regarding the phonological representation). Mapping the incoming auditory or orthographic pattern of any one of these morphological forms to this representation, regardless of whether there is a complete match between input and stored information, constitutes lexical access. What we investigate here is the effect of stem vowel alternations and their representation in the orthography on the difficulty and speed of this process.

The English examples we used above stem from irregular past tense forms, which have been discussed extensively in the English past tense debate (e.g. Clahsen, Eisenbeiss, & Sonnenstuhl-Henning, 1997; Marslen-Wilson & Tyler, 1998; McClelland & Patterson, 2002; Pinker, 1991). Most investigations into the processing of regulars vs. irregulars fall short of explaining the role of either phonology or orthography per se due to a focus on dual vs. single route models to handle systematic vs. idiosyncratic changes and therefore do not address the interacting roles of phonology and orthography. Crepaldi, Rastle, Coltheart, and Nickels (2010), however, reported successful priming with irregular past tense forms in a masked priming experiment (i.e. with visually presented primes), in line with findings reported by Meunier and Marslen-Wilson (2004) for French inflections (but see Kielar, Joanisse, & Hare, 2008; for conflicting results; and cf. Pastizzo & Feldman, 2002, who report a lack of priming in forms such as *taught* ~ *teach* which they attribute to a lack of orthographic overlap).

Importantly, however, vowel alternations occur in regular contexts across many languages, and experiments addressing irregulars cannot answer the question of how such predictable processes are dealt with. Testing the relative contribution of phonology and orthography systematically is even more challenging as it is difficult to find a large enough set of words which behave completely systematically with regard to both aspects.

Here, we use verb inflection in Bengali to approach this question, where a highly regular process of vowel raising which leads to phonological alternations, and a systematic difference in whether or not these vowel changes are represented in the orthography coincide. In other words, Bengali presents the ideal test case for the interacting roles of phonology and orthography in the resolution of inflected forms. The Bengali verb paradigm used in this study provides three distinct categories of pairs of stem and affixed items: those without a phonological

change which are entirely transparent (phonology matches orthography), those which only differ phonologically (with identical orthography) and those which differ phonologically for which this difference is encoded in the writing. This provides the perfect opportunity to investigate the relative contribution of phonology and orthography to the process of resolving an auditory form that is not identical to its stored lexical representation – a combination of factors that are elsewhere hard to find.

Across languages, verb inflection uses forms that differ with respect to the phonology of the stem (e.g. *get* [gɛt] ~ *got*[gɔt]), that involve the addition of an affix (e.g. *pack* [pæk] ~ *pack-ed* [pæk-t]), or both (e.g. *keep* [kiːp] ~ *kep-t* [kɛp-t]), as well as more irregular forms. The inflected forms we investigate in this paper are Bengali present tense forms. Here, suffixes [-i] and [-e] are used to produce the 1ˢᵗ and 3ʳᵈ ᴘᴇʀsᴏɴ sɪɴɢᴜʟᴀʀ, respectively, and adding the [-i] suffix involves raising the stem vowel by one step for most forms (i.e. a regressive assimilation process that adjusts the stem vowel height towards that of the suffix). An example is [lekʰ-e] 'write-3ᴘ.ᴘʀᴇs'[3] with its 1ˢᵗ ᴘᴇʀsᴏɴ form [likʰ-i] 'write-1ᴘ.ᴘʀᴇs', in which the base vowel [e] in [lekʰ-] is raised to [i] (i.e. the vowel height is increased by one step) because the suffix [-i] follows. Corresponding processes occur for base vowels [o] (raised to [u]), [æ] (raised to [e]) and [ɔ] (raised to [o]). This results in a paradigm where there are changes to the stem vowel which are governed by phonology and are entirely predictable. The phenomenon under investigation here concerns all monosyllabic verbs in Bengali, except those with stem vowel [a], which is not raised. Such stem vowel changes due to regressive vowel raising are a common occurrence across languages where a high vowel affix (e.g. [-i]) results in a step-wise raising of the stem vowel (e. g., German *geb-en* [geb-ən] 'give-ɪɴꜰ' ~ *gib-st* [gib-st]'give-2ᴘ.sɢ', where *gibst* is derived via vowel raising from Old High German *geb-ist > gib-ist*; Hock, 1991, p. 212). This phenomenon is especially common diachronically but also well-attested synchronically, and these sound changes are not always represented orthographically (see above).

In the past (Middle Bengali, 1200–1800; Chatterji, 1926/1978), step-wise raising of the stem vowel (low-to-mid, mid-to-high) in the context of high vowel suffixes ([-i, -u]), was an entirely transparent process. Vowels changed in height, while the place of articulation remained the same: [æ, ɔ] became [e, o] and in turn [e, o] became [i, u]. The formal writing system (Bengali [ʃuɖʰːo bʰaʃa] 'pure language'), however, in some cases still reflects the older stage of the language (i.e. does not use different graphemes for the raised vowels). In Modern Bengali, there is a one-to-one correspondence between sounds and letters for the high vowels [i] and [u] and the low vowel [a] is also entirely transparent. By contrast, the low vowels [æ] and [ɔ] differ in terms of their sound-letter correspondence. Thus, Modern Bengali has a system where entirely regular phonological alternations in the verb paradigm differ in terms of the degree to which these changes are represented orthographically: [lekʰ-e] ~ [likʰ-i] 'write-3ᴘ/1ᴘ' লেখে ~ লিখি (distinct spelling of the stem vowel) vs. [kʰæl-e] ~ [kʰel-i] 'play-3ᴘ/1ᴘ' খেলে ~ খেলি (identical spelling of the stem vowel, despite a change in pronunciation).[4] The details of these alternations are given in Fig. 1. Consequently, the recognition of these inflected forms may be affected by multiple, possibly interacting, cues.

While Bengali provides an ideal test case for the interaction of phonological and orthographic information, the phenomenon of vowel alternations is, as outlined above, common across languages and thus the findings presented here are more widely applicable. In the present paper, we approach the question of how listeners identify these inflected forms using two priming paradigms (cross-modal lexical decision in

---

[3] Abbreviations: ᴘʀᴇs = present tense, 1ᴘ = first person, 3ᴘ = third person, ɪɴꜰ = infinitive, sɢ = singular.

[4] In the Bengali writing system, within a syllable the vowel nucleus is a diacritic on the onset consonant. Note that graphemes are not placed linearly in a left-to-right fashion. E.g. for [kʰæ] খে, the diacritic precedes the consonant, whereas for [mu] মু it is attached at the bottom.
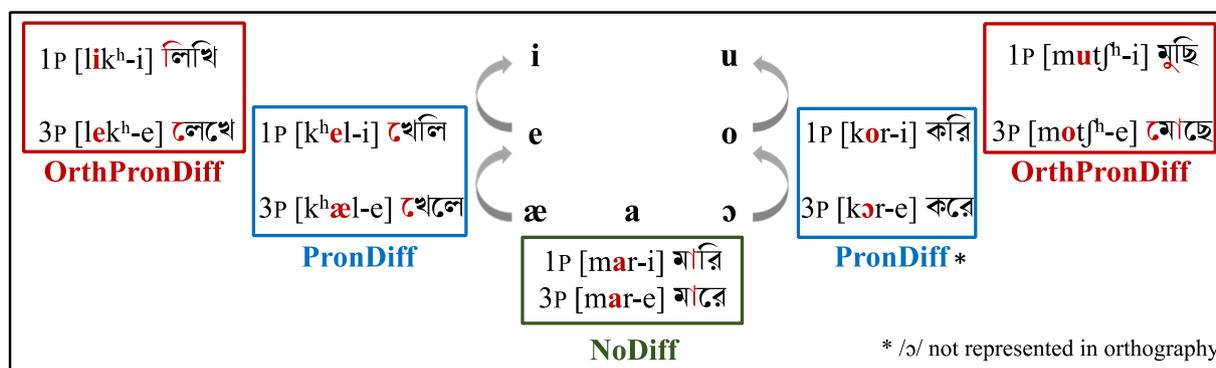
Fig. 1. Examples and explanation of the vowel alternations in 1ᴘ and 3ᴘ Bengali present tense verbs.

Experiment 1, and fragment completion with eye tracking in Experiment 2) with auditory primes. As there is no previous research on the processing of inflected forms with vowel raising, and very little on the interaction of phonological and orthographic information in the processing of inflection, we draw on the literature on orthographic activation in auditory processing, as well as relevant studies of the processing of morphophonological alternations, to provide the background for our investigation. We focus in particular on the activation of orthographic representations when hearing a word, and mechanisms involved in this process.

1.1. Orthographic activation in auditory processing

Psycholinguistic and neurolinguistic investigations, starting with Seidenberg and Tanenhaus' seminal work (Seidenberg & Tanenhaus, 1979), provide convincing evidence for the mutual activation of orthography and phonology in language processing. Using a number of different methods and paradigms, studies show that orthographic information is activated during spoken word processing (Cutler & Davis, 2012; Seidenberg & Tanenhaus, 1979; Taft, 2011; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008; Tanenhaus, Flanigan, & Seidenberg, 1980). While this cross-modality activation seems well established, the precise mechanisms remain under debate. Many accounts point towards this process being automatic (cf. Taft, 2011, for a review), including studies investigating the automaticity of orthographic activation using orthographic congruency effects in priming (e.g. testing whether *dream-gleam* leads to stronger priming than *scheme-gleam*, Chéreau, Gaskell, & Dumay, 2007; but see Pattamadilok, Kolinsky, Ventura, Radeau, & Morais, 2007 for contrasting results). However, whether the locus of the effect is at the lexical or sublexical level is an unresolved question, because experimental techniques that reveal an orthographic consistency effect with existing words (e.g. Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Perre, Pattamadilok, Montant, & Ziegler, 2009; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler & Ferrand, 1998; Ziegler, Petrova, & Ferrand, 2008) have sometimes extended to pseudowords (Pattamadilok, Morais, et al., 2007) and sometimes not (Ventura et al., 2004; Ziegler & Ferrand, 1998). Here, the question is whether words with rimes that are associated with a single spelling (e.g. sl*ing*) lead to faster lexical decision compared to words with rimes that could be spelled multiple ways (e.g. br*ief*, whose rime could also be spelled <-eef> or <-eaf>).

Taft (2011) illustrates how previous findings can all be explained by an Interactive Activation-style model (Taft, 1991) which adds an orthographic component to the TRACE model (McClelland & Elman, 1986). However, most previous experimental results are also compatible with the theory that phonological representations themselves are modified as orthography is acquired. In this account, hearing a word does not necessarily activate a separate orthographic level of representation but the activated phonological representation is different from

one in an illiterate listener. This theory is supported by Perre et al. (2009) who suggest that their ERP results on the consistency effect indicate an involvement of areas associated with phonology, and not of areas such as the Visual Word Form Area (see below). Of relevance for the present work is particularly the automaticity of orthographic activation for existing lexical items: we can assume that perceiving an auditorily presented prime word or prime fragment will not just activate the lexical entry itself, but also an orthographic representation.

Further evidence comes from fMRI studies investigating how learning to read changes the brain; both in the context of comparing illiterate and literate adults and children before and after beginning to read. One area of interest here is the Planum Temporale (PT), which represents phonemic categories (Chang et al., 2010; Mesgarani, Cheung, Johnson, & Chang, 2014) and was shown to be much more active in literates compared to illiterates while listening to speech (Dehaene et al., 2010). PT activation is also correlated with tasks relating to phonemic awareness. A second area under investigation is the Visual Word Form Area (VWFA; Dehaene et al., 2010; Dehaene & Cohen, 2011). While this area is generally understood to be inactive during listening (e.g. passive sentence comprehension), Dehaene et al. (2010) found activation of the VWFA during auditory lexical decision tasks. They proposed that the VWFA may be particularly involved where the orthographic code needs to be activated in order to make a difficult decision, such as lexical decision or rhyming – in particular in the presence of a spelling conflict (e. g. *pint/mint*). Dehaene et al. (2010, 2015) interpret this as top-down activation of the orthographic form.

1.2. Effects of morphophonological alternations on lexical access

While no studies to date have specifically addressed the effect of orthography in the processing of regular inflectional paradigms, there has been research on whether morphophonological alternations cause delays in lexical access. One such example uses vowel alternations in German morphologically complex forms between back and front vowels in words such as *Stock* 'stick-SG' ~ *Stöck-e.* 'stick-PL'. Scharinger, Lahiri, and Eulitz (2010) investigated the processing of such forms in contrast to non-alternating *Stoff* 'fabric-SG' ~ *Stoff-e* 'fabric-PL' using EEG, finding stronger mismatch negativity if the non-alternating base form (*Stoff*) was preceded by a stem exhibiting the fronted vowel. Phonological alternations, however, need not impede processing. In behavioural studies on derivational morphology equal facilitation has been observed for derivations with phonological alternations and those with phonological and orthographic changes (e.g. Fowler et al., 1985; Stanners, Neiser, Hernon, & Hall, 1979). Marslen-Wilson et al. (1994) investigated derivational morphology in a series of cross-modal lexical decision tasks and concluded that phonological alternation of suffixed items (e.g. *elusive ~ elude* or *serenity ~ serene*) had no detrimental effects on the facilitation of the base item as long as they were equally semantically transparent. Thus, for instance, *van-ity* [væn-ɪti] primes *vain* [veɪn] just

as well as *happiness* [hæpi-nəs] primes *happy* [hæpi].

Lahiri and Reetz (2010) examined a morphophonological vowel alternation similar to that in the present study using German singular and plural nouns. The stem vowel in German singular nouns with rounded back vowels changes to an umlauted vowel in the plural (e.g. *Sohn* [zoːn] 'son-sɢ' ~ *Söhne* [zøːne] 'son-ᴘʟ') with only very few exceptions (e.g. *Boot* [boːt] 'boat-sɢ' ~ *Boote* [boːte] 'boat-ᴘʟ') where this is not the case. In an auditory delayed priming task (with several items intervening between prime and target), they used the plural forms (e.g. [zøːne] and [boːte]) to prime singular targets (e.g. [zoːn] and [boːt]) to determine whether the plurals with the umlauted base vowel resulted in a difference in the degree of activation of the singular target compared to those cases where no umlauting took place. Their results show no difference in the degree of priming between the two conditions and they thus conclude that the plurals showing the vowel alternation are equally effective at activating the lexical entry as those which do not, regardless of the difference in the surface forms.

Previous research on the processing of morphologically related regularly inflected words has shown that the base is always activated when an inflected form is encountered and that inflection priming seems to generate stronger facilitation than derivational priming (see Amenta & Crepaldi, 2012 and Leminen, Smolka, Duñabeitia, & Pliatsikas, 2019 for reviews). Research on verb paradigms has focused on the discussion of possible differences in representation and access between regular and irregular forms (e.g. Clahsen et al., 1997; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Marslen-Wilson & Tyler, 1998; McClelland & Patterson, 2002; Pinker, 1991) with some accounts proposing a difference in terms of storage and therefore access. According to those models, while irregular items are claimed to be stored as separate lexical entries, for regular inflection the base is stored and the inflected forms are decomposed or constructed rather than stored and retrieved independently. Connectionist models, by contrast, propose a single-mechanism framework for both types of word forms (McClelland & Patterson, 2002). Several studies focused specifically on the question whether irregular inflections prime the corresponding base form. Meunier and Marslen-Wilson (2004) used regular and irregular French inflections, some of which included vowel alternations (e.g. regular: *aimerons* [ɛmrɔ̃] vs. *aimons* [ɛmɔ̃] to prime the target *aimer* [ɛme:] and irregular with vowel change: *levont* [ləvɔ̃] vs. *lèvent* [lɛv] to prime the target *lever* [ləve:]) Their results indicated equal amounts of priming for both the regular and irregular forms, both in cross-modal repetition priming and masked priming, which eliminates semantic effects. For English, Pastizzo and Feldman (2002) reported priming effects in visual masked priming with past tense forms involving only a vowel change (*fell* ~ *fall*), but no corresponding effects for forms such as *taught* ~ *teach*. This was attributed to the lack of orthographic overlap. However, Crepaldi et al. (2010) showed that consistent priming effects emerge once the orthographic overlap calculation is not based on a purely slot-based metric but uses spatial coding, where grapheme identity is weighted more heavily compared to position within a string (Davis & Bowers, 2006). Similar to Pastizzo and Feldman (2002), Crepaldi et al. (2010) compared orthographically and morphologically related primes (*fell* ~ *fall*) against (a) orthographically but not morphologically related primes (*full* ~ *fall*) and (b) neither orthographically nor morphologically related primes (*hope* ~ *fall*). Importantly, however, orthographical overlap was calculated using both left-aligned slot coding (McClelland & Elman, 1986) and spatial coding (Davis & Bowers, 2006), eliminating the shortcomings of Pastizzo and Feldman's (2002) method of accounting for orthographic overlap.

Studies on irregular inflection paradigms aside, the transparency of the phonological form and the interaction of phonology and orthography have not yet been thoroughly investigated. The verb paradigm we investigate in this paper is entirely regular and the phonological alternations described above are fully predictable. Therefore, as the lexical entry is activated, listeners can draw on a phonological rule which is entirely consistent across verbs. We assume all inflected forms will

activate one lexical representation, that of the stem, and our investigation is thus focused on the more fine-grained differences across different groups of words, as well as on the time course of processing. As we shall see, while a lexical decision paradigm is not sensitive enough to distinguish between words with different degrees of orthographic and phonological change, the time course of eye movements during discrimination of similar forms (rather than just lexical access) provides deeper insight into the processes involved.

1.3. The present study

The current paper investigates the relationship between phonology and orthography from a different angle compared to previous work. In the present work, stimuli are inflected verb forms of the same stem which are subject to a particular type of vowel alternation: raising the vowel one step in the context of a high vowel suffix [-i, -u] in Bengali verb paradigms. In this paper, we restrict ourselves to Standard Colloquial Bengali as spoken in Kolkata, India (Chatterji, 1926/1978). In Bengali, person marking in the present tense is indicated by a final vowel: [-i] for 1ˢᵗ ᴘᴇʀsᴏɴ ([kʰel-i] 'play-1ᴘ.ᴘʀᴇs') and [-e] for 3ʳᵈ ᴘᴇʀsᴏɴ ([kʰæl-e] 'play-3ᴘ.ᴘʀᴇs'). The citation form of verbs (verbal noun) ends with [-a] ([kʰæl-a] 'play-ᴠɴ')[5]. Certain combinations of vowels trigger vowel raising. The high vowel [-i] of the 1ˢᵗ ᴘᴇʀsᴏɴ suffix will trigger the stem vowels [æ, e, ɔ, o] to raise to [e, i, o, u], while the 3ʳᵈ ᴘᴇʀsᴏɴ suffix [-e] does not have any effect. In addition, the stem alternations [e ~ i] and [o ~ u] are reflected in orthography while [æ ~ e] and [ɔ ~ o] are not. In this study, we investigate three groups of verbs with stem vowels [a, æ, ɔ, e, o] (see Fig. 1 for examples):

- *NoDiff*: low [a] does not change; no difference in phonology or orthography
- *PronDiff*: low [æ] and [ɔ] are raised to mid [e] and [o]; difference only in phonology which is not reflected in the orthography
- *OrthPronDiff*: mid [e] and [o] are raised to [i] and [u] respectively; difference in both phonology and orthography

Verbs in the *NoDiff* condition are entirely transparent due to the lack of change in the stem vowel [a] and the changes in the *OrthPronDiff* condition are equally unambiguous as the vowel alternation is reliably represented in the orthography of these items in both front and back vowels. In the alternation of the back vowels [o]-[u] with the root vowel [o], both forms are transparent in their letter-sound correspondences and both vowels are explicitly marked (e.g. [mo] গো vs. [mu] মূ for [motʰ-a] 'wipe-ᴠɴ'). In the cases of [e]-[i] alternation with the root vowel [e], both forms have their respective digraphs matching the pronunciation (e.g. [le] লে vs. [li] লি in [lekʰ-a] 'write-ᴠɴ').

Where the *PronDiff* condition is concerned, however, the relationship between phonology and orthography is less transparent and differs between front and back vowels. In the alternation of the front vowels [æ]-[e] with the root vowel [æ], the vowel [æ] in the 3ʳᵈ ᴘᴇʀsᴏɴ is written with the same digraph as that for [e]. Thus [kʰæ] and [kʰe] are both represented orthographically by the digraph খে for the verb [kʰæl-a] 'play-ᴠɴ'. Note that this involves the same grapheme used to represent [e] in the [e]-[i] alternation above. The representation of the alternation of the back vowels [ɔ]-[o] with the root vowel [ɔ] is complicated by the fact that the vowel [ɔ] has no independent grapheme in Bengali. Thus, the bare consonant indicates the vowel [ɔ]. Within this paradigm, the pronunciation of the raised vowel [o] is not reflected in writing, unlike in the [o]-[u] alternation above, and thus both [kɔ] and [ko] are represented only by the grapheme for [k] ক as in করা [kɔr-a] 'do-ᴠɴ'. Consequently, while in neither case the changes are represented orthographically, in the [æ]-[e] paradigm, the written vowel is the one for [e]

---

[5] Abbreviations: ᴠɴ = ᴠᴇʀʙᴀʟ ɴᴏᴜɴ. The verbal noun (root + -a) is the citation form in Bengali.

while in [ɔ]-[o] there is no overt digraph, which is usually interpreted as [ɔ].

This pattern results in different degrees of transparency in terms of (a) the correspondence between sounds and letters and (b) between sounds and base form. With regard to the question of whether the first syllable vowel predicts the suffix (and hence the *person* attribute of the word in question), this may result in ambiguity. If [i] or [u] are heard in the context of this verb paradigm, both are an unambiguous indicator of a 1st PERSON form from the *OrthPronDiff* group and both vowels are transparently represented in writing. [æ] and [ɔ] must be a base form from the *PronDiff* group, but will not come with distinct orthography. In the case of [a], [e] and [o] the phonology does not indicate unambiguously whether this will become a first or third person form or verbal noun. As the low vowel [a] does not change, it occurs in all forms of the verb. It is therefore at least clear that the verb in question belongs to the *NoDiff* group (although the *person* attribute can only be resolved once the suffix is perceived). The vowels [e] and [o], however, can either be the base vowel (3rd PERSON or VN), i.e. the verb belongs to the *OrthPronDiff* group, or they could be the result of raising from [æ] and [ɔ], i.e. the verb belongs to the *PronDiff* group. Thus, in addition to the opacity in the sound-letter correspondence, these vowels are also ambiguous in their indication of the paradigm.

We assume that recognising a word form completely after hearing auditory input involves the activation of a lexical representation which consists of the stem and information of all possible morphological affixes and their phonological exponents (e.g. all inflectional forms of the verb and the phonological alternations in stem vowels in our case).

Using a cross-modal lexical decision task and a fragment-completion eye-tracking study, we ask three central questions:

(a) Does the difference in stem vowel resulting from vowel raising affect access to the mental representation?

(b) Does reflection of the difference in the orthography affect access to the mental representation?

(c) Is the ability of the stem vowel to uniquely predict the suffix (or not) the factor that drives the speed of form discrimination (Exp. 2)?

We first present a primed lexical decision task, where the auditory primes are either 1st or 3rd PERSON forms and the visual targets are the verbal noun (base form; see Table 1). The aim is to establish whether forms that differ in terms of the stem vowel, or in terms of the stem vowel and spelling, are less effective at priming the corresponding targets, i.e. whether these forms take longer to activate the lexical entry, and whether differences emerge for 1st vs. 3rd PERSON primes. This study employs a cross-modal paradigm to avoid effects based on auditory or visual memory traces as well as to allow immediate presentation of the target at the offset of the auditory prime to minimise effects of strategic processing (e.g. McQueen & Sereno, 2005). We expect facilitation of the word/non-word decision in all conditions since primes are always semantically identical and semantic priming effects have been shown to cause the greatest priming effects, especially when the inter-stimulus-interval (ISI) is short (e.g. Bentin & Feldman, 1990). A control condition with unrelated primes is therefore not necessary and was not included in this design. As the 3rd PERSON shares the stem vowel with the

verbal noun (target), there is a greater degree of overlap between these two forms across all three word groups and the facilitation effect of overlap in semantics, morphology and form may be additive. We assume the underlying cause of this facilitation to be a result of repeated automatic activation of the same pre-lexical representation (see McQueen & Sereno, 2005). All primes will activate the same prelexical representation shared by the verbal noun and thus result in faster access. However, if the difference between the stem vowel in the 1st PERSON primes and their targets affects the ease of access to the lexical entry for the stem, 3rd PERSON primes may show greater facilitation of the verbal noun target than 1st PERSON primes in both the *PronDiff* and *OrthPronDiff* conditions since the 3rd PERSON primes' first segment is identical to that of the target. In addition, if the difference in orthographic representation of the vowel alternation plays a role in processing, we would expect to see a difference between the *OrthPronDiff* condition and the *PronDiff* condition as greater overlap may result in faster access and therefore a greater amount of pre-activation of the verbal noun target. Should the vowel alternation alone not affect access, this may even lead to no difference emerging between 1st PERSON and 3rd PERSON in the *PronDiff* condition. In the *NoDiff* condition, no difference in the facilitation resulting from the two primes is expected because the stem vowels are identical. While this primed lexical decision task is a standard way of assessing lexical access, one caveat is that differences between the primes may in this case not be sufficiently strong to attenuate or amplify the semantic priming effect that should arise in all cases, since each prime and target pair is semantically identical in all three conditions. In addition, the 3rd PERSON is frequently considered a default form (e.g. Kiparsky & Tonhauser, 2013, p. 2077; Bybee, 2001; Lahiri, 1982) which may also lead to faster activation of the stem after 3rd PERSON primes across all conditions that could override any effect of condition as this access route may be more deeply entrenched.

In order to investigate the precise time course of processing in the different conditions, we set up a second experiment, this time using eye tracking. By contrast to the priming task, the eye tracking experiment (Exp 2) required subjects to make an active decision between two potential forms. In other words, in addition to lexical activation a precise analysis of the morphological form is required. In a cross-modal fragment completion task, subjects were presented with a first-syllable fragment of either a 1st PERSON or a 3rd PERSON form in the *NoDiff*, *PronDiff* or *OrthPronDiff* condition. Presentation of the auditory fragment (e. g. [kʰæ-] from [kʰæle]) was immediately followed by a visual display showing the full orthographic forms of both 1st and 3rd PERSON of the same verb (e.g. খেলি [kʰeli] and খেলে [kʰæle]). Subjects were instructed to press a button corresponding to the match, and their eye movements were tracked for a two-second window. Since this task requires precise discrimination between 1st and 3rd PERSON, we hypothesized that eye movements would provide fine-grained insight into the time course of the matching process. Using growth-curve analysis (Mirman, 2014) we can construct a detailed model capturing the time course of looking across the trial. This involves fitting polynomials to the curve shapes corresponding to the likelihood of participants looking at a target on the screen, here, the word (1st PERSON form or 3rd PERSON form) which they first fixate. Depending on whether subjects perceive this word as a match for the auditory input they heard, or not, they will linger on this item or move away towards the remaining item. In particular the comparison of looking patterns on trials where subjects encounter the matching item first with those on trials where the non-matching item is encountered first, allows us to determine whether (and at what point in time) participants discriminate between the two forms: if the curves are identical, subjects are not able to discriminate between the forms; if the curves diverge (lingering on the match, moving away from the non-match), then subjects are able to discriminate. In addition, as we shall see, the detailed eye movement patterns do not just provide a binary answer to the discrimination question, but also provide insight into the relative difficulty of the decision.

Table 1
Example stimuli for Experiment 1.

Condition	NoDiff low vowel/ a/		PronDiff low-to-mid		OrthPronDiff mid-to-high	
	3P	1P	3P	1P	3P	1P
Prime	mar-e	mar-i	kʰæl-e	kʰel-i	lekʰ-e	likʰ-i
Target	mar-a		kʰæl-a		lekʰ-a	
	মারা		খেলা		লেখা	
	'hit-VN'		'play-VN'		'write-VN'	

2. Experiment 1

We conducted a cross-modal lexical decision task (with auditory primes and visual targets) to investigate whether there is a difference between the degree of facilitation of a verbal noun target after 1st vs. 3rd PERSON primes depending on the degree of overlap of prime and target in both phonological and orthographic information. When the listener hears the prime, the lexical entry is activated, including its orthography, and matched to that of the target. This match can either be a complete match or a partial match with varying degrees of phonological and orthographic overlap, which may result in different patterns of facilitation.

2.1. Methods and design

2.1.1. Participants

Thirty-four native Bengali speakers took part in the study (average age: 19.2 years) and all had normal or corrected-to-normal vision and did not report any language disorders such as dyslexia or any hearing deficits. Participants were students at Satyapriya Roy College of Education, Kolkata (India) where the experiment was conducted.

2.1.2. Stimuli and design

Targets were always the verbal noun (with the suffix [-a]) of a particular verb (presented visually) while the primes were either the 1st or 3rd PERSON singular present tense verb (presented auditorily). As outlined above, there are three different possibilities for the phonological and orthographic relationship between primes and targets due to a regressive vowel raising rule and the orthographic representation, which led to the three experimental conditions (cf. Table 1 for examples and Appendix A/Table 5, Table 6 and Table 7 for full stimulus lists).

Stimuli were divided into two lists to allow for priming with either the 1st PERSON or 3rd PERSON inflected form while ensuring each participant was only exposed to any given target once throughout the experiment. The number of first and third person verbs in each condition was counterbalanced across the two lists. Trials were pseudo-randomized within each list and care was taken to avoid the repetition of the same condition in adjacent trials. Each of the three conditions consisted of 24 verbal nouns paired once with the 1st and once with the 3rd PERSON prime. In addition, three sets of 24 non-word targets with real word primes were constructed using the same pattern of verbal noun targets and inflected primes. Primes for non-word targets were real-word 1st PERSON or 3rd PERSON forms and non-word targets were constructed to show overlap with the corresponding prime and to end in the verbal noun suffix -a (e.g. primes /dʒal-i/ 'burn-1P.PRES' or /dʒal-e/ 'play-3P.PRES' for target */dʒad-a/).

The inter-stimulus-interval between prime and target was 0 ms meaning the visual target was presented at the offset of the auditory prime. The targets were presented for 1000 ms followed by an inter-trial-interval of 1500 ms when a bleep indicated the start of the next trial.

2.1.3. Procedure

The experiment was conducted in a quiet and darkened room at Satyapriya Roy College of Education in Kolkata, India. Participants were tested in groups of sixteen. Targets were projected onto a screen and auditory primes were played through individual closed-ear headphones (SONY MDR110 LP). Each subject responded to the visual stimuli via individual custom-made two-button boxes. Participants used their dominant hand to indicate a 'yes' response. Response data and reaction times were collected with custom-made experimental hardware and software (Reetz & Kleinmann, 2003).

Before each session participants were instructed (verbally) to respond to visual items as quickly and accurately as possible and to not respond to or pay any particular attention to what they hear (auditory primes). Following the instructions, participants were given the

opportunity to ask questions and written consent was obtained. After this a ten-item practice task was presented to familiarise subjects with the experiment and this task was immediately scored and repeated until the experimenters were satisfied that all participants had understood the task correctly. Once all subjects were comfortable with the task, the main experiment was run in two blocks of approximately twelve minutes with a short break in between. Smaller breaks, indicated by a countdown from five to one on the screen, were built into each block after every 16 items.

2.2. Results

2.2.1. Data cleaning procedure

All subjects and targets with error rates greater than 30% were excluded from the analysis (one subject (3.30% of total trials) and eight targets (2.57%) removed). Thirty-three subjects remain for the final analysis. All reaction times (RT) outside $+/-$ 2 standard deviations of the individual subject mean were also excluded (2.02%). Reaction time data was log transformed based on a Box-Cox transformation (Box & Cox 1964; -0.1010) and visual inspection of Q-Q plots and histograms.

2.2.2. Reaction time analysis

We performed a linear mixed model analysis in R (using the lme4 package; Bates, Maechler, Bolker, & Walker, 2015) with *Condition* and *Person* (1st PERSON or 3rd PERSON prime) as fixed effects and *Target* and *Subject* as random factors (see Table 2 for mean RTs). A model with random intercepts only[6] provided the best fit, as the additions of random slopes either resulted in models which did not converge or produced indicators of overfitting. In addition, model fit and distribution of residuals were visually inspected using Q-Q plots and histograms. The results show no overall effect of *Condition* ($\chi^2(4) = 4.444$, $p = .349$) or of the interaction *Condition x Person* ($\chi^2(2) = 0.8816$, $p = .644$) while there is a significant effect of *Person* ($\chi^2(3) = 8.7342$, $p = .033$) indicating faster response latencies after 3rd PERSON primes than 1st PERSON primes. As there is no significant interaction, the differences in the degree of overlap do not seem to directly affect reaction times.

2.2.3. Error analysis

The data shows very low error rates across all conditions (cf. Table 2) and was analysed in R using a binomial logistic regression[7] with *Condition* and *Person* as fixed factors and error as the dependent variable. The analysis shows that neither the addition of *Person* ($\chi^2(1) = 1.139$, $p = .286$) nor that of *Condition* ($\chi^2(2) = 0.464$, $p = .793$) significantly improves model fit, and the interaction of *Person x Condition* ($\chi^2(2) = 0.268$, $p = .875$) did not reach significance either. Thus, no further investigations of differences between conditions were carried out.

2.3. Discussion

The analyses show a main effect of *Person* (1st PERSON vs. 3rd PERSON)

Table 2
Experiment 1 results for RT (in ms) and errors (in %).

Condition	RT (ms)		Errors (%)	
	3rdP	1stP	3rdP	1stP
NoDiff	591	599	4.88	5.32
PronDiff	603	616	4.99	6.78
PronOrthDiff	610	625	5.15	6.13

---

[6] lmer(LogRT~Condition*Person+ (1|Subject) + (1|Target), data = VR_January)

[7] fit <− glm(Correct~(Relatedness+Condition)^2, data = VR_January_Errors, family = 'binomial')

but no effect of *Condition* and no statistically significant interaction. Thus, it seems that, while overall 3rd PERSON primes generate faster responses to the target, this pattern is not affected by the degrees of difference in phonological and orthographic representation but is likely either an effect of the greater shared material between the 3rd PERSON primes and their targets as the stem of the 3rd PERSON prime is identical to that of the verbal noun target or of the greater frequency of use of the 3rd PERSON form or, indeed, a combination of those factors. As semantic information has been shown to play a very prominent role in cross-modal lexical decision tasks (e.g. Amenta & Crepaldi, 2012; Marslen-Wilson et al., 1994), these results are not altogether surprising and it seems that, if there is a difference to be found, a different methodological approach is necessary.

3. Experiment 2

In Experiment 2 we approached the question of lexical access with cross-modal fragment completion. The aim was to obtain more precise information about the time course of word recognition across the different conditions. Eye tracking was therefore used in order to gain fine-grained insight into mental processes. This task involved listening to a first-syllable fragment and choosing the matching visual word form (entire word in Bengali script) from two alternatives. First-syllable fragments stemmed from the 1st PERSON or 3rd PERSON form, and both of these forms were presented visually as potential targets (match and non-match). Participants were asked to respond by pressing a button, and their eye movements were tracked throughout the trial. The idea here is that rather than being a potential hindrance due to a mismatch between auditory form and citation form the vowel alternation could provide an advantage because the upcoming suffix may be predictable on the basis of just the first syllable fragment. The question is whether listeners can utilise this information, and if so whether a representation of the stem vowel change in orthography matters. Our prediction of the processes involved is as follows: Upon hearing the first syllable fragment, the listener needs to activate a lexical entry, which will retrieve paradigm information, i.e. in particular different sets of morphological features together with their phonological realisation. The best match for the auditory fragment is selected, thereby determining the morphological form. We hypothesized further that due to the experience with alternating stem vowels in the *OrthPronDiff* condition being realised with distinct graphemes, and alternating stem vowels in the *PronDiff* condition being realised with identical graphemes, the process of lexical activation and morphological resolution would be easier, or faster, for the *OrthPronDiff* condition.

The rationale for our eye tracking study was that the time course of eye movement patterns would reveal the difficulty of deciding between the two competing forms, and thereby the speed of fragment completion or full morphological resolution. In particular the time course of looking incorporates the time participants take to inspect each item they encounter in turn (match or non-match).

For the *NoDiff* condition the task is undecidable by definition: both visual items are potential matches for the fragment, since the stem vowel in this condition does not change (note that for the purpose of the analysis one target item was still designated the nominal "match"). Any differences we observe here between different trials should stem from biased responding to 1st vs. 3rd PERSON.

For the *PronDiff* and *OrthPronDiff* conditions, by contrast, the task ought to be decidable. Depending on whether orthography plays a role we may expect differences between these conditions. In the *PronDiff* condition, the stem vowels in both target forms are represented by the same graphemes, so an orthographic representation activated upon hearing the first syllable cannot aid discrimination. In the *OrthPronDiff* condition, however, such an orthographic representation could make discrimination easier as overall activation patterns should be more distinct between a 1st and 3rd PERSON form.

Fragment completion further essentially involves the prediction of

the suffix of the word. Let us consider exactly how listeners could do this in the *PronDiff* and *OrthPronDiff* conditions. The full lexical form could be accessed as we predicted above, by activating a lexical item, which retrieves paradigm information, i.e. activates a set of corresponding morphological forms, one of which will match best, or alternatively by using just the stem vowel to predict the suffix vowel. Due to the stepwise raising system, mid stem vowels [e], [o] can occur both in a base form and in a raised form, and therefore on their own do not predict the suffix unambiguously, whereas low ([æ] or [ɔ]) and high ([i] and [u]) vowels only occur either in a base *or* in a raised form and therefore predict the suffix unambiguously (low vowels must be followed by [-e], i.e. the 3P suffix, high vowels must be followed by [-i], the 1P suffix), see Table 3.

As a consequence, there are two competing hypotheses about solving the fragment completion task. In both cases we hypothesize that orthography plays a role. Under the first hypothesis the stem vowel identity, and therefore its ambiguity or lack thereof, is crucial. Under the second one the stem vowel identity does not affect processing much, and instead the paradigm information that comes with activating a lexical entry upon hearing the target fragment plays an overriding role. This leads to two distinct hierarchies in the difficulty of discrimination, depending on which hypothesis holds. Under the first hypothesis (stem vowel ambiguity is important), we expect the following (easy to difficult):

Hierarchy 1 (stem vowel ambiguity is important).

(1) 1P OrthPronDiff: [li-] from [lik^h-i] – *unambiguous stem vowel, transparent orthography.*
(2) 3P PronDiff: [k^hæ-] from [k^hæl-e] – *unambiguous stem vowel, nontransparent orthography.*
(3) 3P OrthPronDiff: [le-] from [lek^h-e] – *ambiguous stem vowel, transparent orthography.*
(4) 1P PronDiff: [k^he-] from [k^hel-i] – *ambiguous stem vowel, nontransparent orthography.*
(5/6) NoDiff: [ma-] from [mar-e] or [mar-i] – *undecidable.*

Under the second hypothesis (stem vowel ambiguity is unimportant) the hierarchy is different, with orthography being the main driver (easy to difficult):

Hierarchy 2 (stem vowel ambiguity is not important).

(1/2) 1P OrthPronDiff: [li-] ([lik^h-i])/3P OrthPronDiff: [le-] ([lek^h-e]) – *transparent orthography.*
(3/4) 1P PronDiff: [k^he-] ([k^hel-i])/3P PronDiff: [k^hæ-] ([k^hæl-e]) – *nontransparent orthography.*
(5/6) NoDiff: [ma-] from [mar-e] or [mar-i] – *undecidable.*

Here we expect that the stem vowel does not play any major role because as soon as the fragment is heard a lexical item is activated, which in turn retrieves the paradigm (i.e. a set of morphological forms). In particular the ambiguity of the stem vowels [e], [o] is therefore irrelevant because potential competing forms that have these stem vowels are no longer active. Here, we expect the question of whether the stem vowels of match/non-match are orthographically distinct to play the biggest role.

Table 3
Stem vowels and possible morphological forms.

Stem vowel	Candidate morphological form(s)
[i]	1P (OrthPronDiff)
[u]	1P (OrthPronDiff)
[e]	3P (OrthPronDiff) or 1P (raised from [æ], PronDiff)
[o]	3P (OrthPronDiff) or 1P (raised from [ɔ], PronDiff)
[æ]	3P (PronDiff)
[ɔ]	3P (PronDiff)
[a]	1P (NoDiff) or 3P (NoDiff)

3.1. Methods and design

The eye tracking experiment conducted in this study is a cross-modal fragment completion task involving an auditory first-syllable fragment presented before two visual targets (match and non-match). Subjects were asked to decide from which word the auditory fragment was taken and respond with a button-press. We assessed both manual responses and eye movement patterns to match vs. non-match items to examine subjects' ability to discriminate the competing morphological forms.

3.1.1. Participants

Twenty-six native speakers of Bengali (average age: 23 years and 6 months) took part in this experiment, which was conducted at Jadavpur University, Kolkata (India). Subjects were mostly students at the university. All subjects had normal or corrected-to-normal vision. Four subjects were excluded from the eye tracking analyses due to below-chance accuracy on the manual responses (see below), leaving 22 subjects and 3168 trials for further analysis.

3.1.2. Stimuli

Words in the *NoDiff*, *PronDiff* and *OrthPronDiff* conditions were the same as in Experiment 1. Stimuli were recorded by a female native speaker of Bengali. Auditory fragments comprised the first syllable of the 1st PERSON or 3rdPERSON form (e.g. [kʰe-] from [kʰel-i], [kʰæ-] from [kʰæl-e]). Fragments were cut so that corresponding forms were equal in duration (mean duration: 241 ms). For the words used in the experiment there are no competing monosyllabic verb stems in Bengali, i.e. forms that begin with the same fragment but belong to a different verb do not exist. The written word forms of corresponding 1st PERSON and 3rd PERSON served as match/non-match pairs (see Table 4).

In total there were 144 trials, half of which contained a 1st PERSON fragment and half of which contained a 3rd PERSON fragment. Each word was used twice, once as a 1st PERSON target and once as a 3rd PERSON target. Note that for the *NoDiff* condition the distinction is purely nominal as the syllables for 1st PERSON and 3rd PERSON fragments are identical and subjects would have no way of knowing which of the visual items was declared the "match". The trial sequence was blocked, so that the first and second half of the experiment contained one occurrence of each word, respectively, with the order of 1st vs. 3rd PERSON counterbalanced across subjects. One third of the stimuli each contained words from the *NoDiff, PronDiff, OrthPronDiff* group of words, respectively.

Targets were presented on screen in a vertical arrangement with one item displayed 160 pixels above and one 160 pixels below a central fixation cross. This spatial configuration meant that the left edges of the two words were approximately equally distant from the central fixation. Half the trials presented the match at the top and the non-match at the bottom; the other half was reversed.

3.1.3. Procedure

Subjects received instructions about the task from a native speaker of Bengali. Written consent was obtained. A chin rest was used to place the subject in a stable position 60 cm in front of a 19-in. monitor. Subjects were instructed to use two buttons on a game pad in order to indicate which item on the screen (top or bottom) corresponded to the fragment they heard. They were asked to guess if they were unable to tell which item the fragment was taken from (i.e. in particular for any items from the *NoDiff* condition, which was, however, not mentioned specifically during instructions). A nine-point calibration was performed prior to the start of the experiment. Twelve practice trials were presented at the beginning of the procedure, before the onset of the experimental trials. Each trial started with a dot presented at the centre of the screen. Once the participant fixated the dot, the experimenter initiated the trial. The auditory fragment was presented through headphones (SONY MDR110) while a central fixation cross was visible on the screen. After 300 ms, the visual targets appeared above and below the fixation cross and remained on screen for 2000 ms. Eye movements were recorded throughout the procedure with an Eye Link 1000 Plus eye tracker.

3.2. Results

3.2.1. Manual responses

We first calculated the accuracy of manual responses over all decidable trials (*PronDiff* and *OrthPronDiff* condition). The average accuracy was $M = 0.67$ ($SE = 0.03$). After excluding four subjects who showed accuracy <0.51, the mean accuracy was $M = 0.72$ ($SE = 0.01$).

A repeated-measures ANOVA with factors *Condition* and *Person* revealed only a main effect of *Condition* ($F(1,21) = 156.39, p < .0001$), with the main effect of *Person* ($F(1,21) = 1.87, p = .186$) and the interaction of *Condition* and *Person* ($F(1,21) = 3.1, p = .093$) failing to reach significance. Responses in the *PronDiff* condition were not significantly different from chance (0.5), $M = 0.54, SE = 0.02, t(21) = 1.62, p = .12$. In the *OrthPronDiff* condition responses reached high levels of accuracy ($M = 0.91, SE = 0.02, t(21) = 23.02, p < .0001$).

However, because the front and back vowels in the *PronDiff* condition are realised differently in Bengali script, with back vowels *not* represented as a grapheme, but front vowels represented by the *same* grapheme, an additional analysis aimed to uncover potential differences in accuracy on the basis of vowel type. Because items in the *NoDiff* condition always contain the low vowel [a], a second ANOVA was conducted on only the response data from *OrthPronDiff* and *PronDiff* conditions, with factors *Condition* (*OrthPronDiff*, *PronDiff*), *Person* (1,3) and *Vowel Type* (front, back). This revealed a main effect of Condition ($F (1,21) = 158.7, p < .0001$) and a significant interaction of *Condition* and *Vowel Type* ($F(1,21) = 10.81, p = .004$). All other main effects and interactions were non-significant ($Fs < 2.57, ps > 0.12$). Follow-up single-sample *t*-tests against chance showed that in the *PronDiff* condition responses for items with back vowels had above chance accuracy ($M = 0.576, SE = 0.032, t(21) = 2.391, p = .026$), whereas responses for items with front vowels were at chance ($M = 0.498, SE = 0.024, t(21) = 0.079, p = .93$).

3.2.2. Overall proportion of looking at the matching item

In a first analysis, we calculated the proportion of looking falling onto the matching item across the entire trial as the total amount of looking falling onto the matching item divided by the total amount of looking falling on either the match or the non-match item (see Fig. 2 for distribution of scores).

These proportion scores were subjected to a repeated-measures ANOVA with factors *Condition* (*NoDiff*, *PronDiff*, *OrthPronDiff*) and *Person* (1,3). This revealed a significant main effect of *Condition* (all Greenhouse-Geisser corrected; $F(1.828, 38.388) = 94.937, p < .001, \eta^2_p = 0.819$). The main effect of *Person* ($F(1,21) = 0.63, p > .43, \eta^2_p = 0.029$) and interaction of *Condition* and *Person* ($F(1.457,30.605) = 1.944, p = .169, \eta^2_p = 0.085$) were not significant. Posthoc tests showed that the *OrthPronDiff* condition ($M = 0.677, SE = 0.014$) resulted in longer target looking than the *PronDiff* condition ($M = 0.523, SE = 0.008, t(21) = 10.88, p < .001$) and the *NoDiff* condition ($M = 0.505, SE = 0.007; t(21) = 11.31, p < .001$). Target looking in the *PronDiff*

Table 4
Example stimuli for Experiment 2.

Condition	NoDiff low vowel/ a/		PronDiff low-to-mid		OrthPronDiff mid-to-high	
	3P	1P	3P	1P	3P	1P
Fragment	ma-	ma-	kʰæ-	kʰe-	le-	li-
Match	mar-e*/mar-i*		kʰæl-e	kʰel-i	lekʰ-e	likʰ-i
	মারে / মারি		খেলে	খেলি	লেখে	লিখি
Non-match			kʰel-i	kʰæl-e	likʰ-i	lekʰ-e

Note. *visual targets representing 1st PERSON and 3rd PERSON forms are equally good matches for first syllables in the NoDiff condition since the first syllable is identical across both forms.
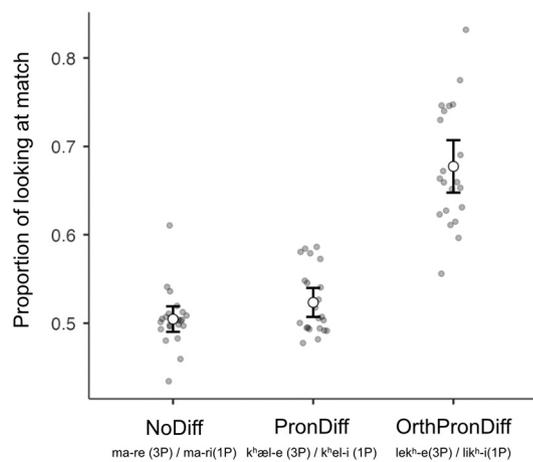
Fig. 2. Looking proportion directed at the matching item across the three conditions.

condition and *NoDiff* condition was not statistically different ($t(21) =$ 1.62, $p = .364$, all Bonferroni-corrected for multiple comparisons).

3.2.3. Detailed eye movement analysis

While the above analysis appears to indicate that responses for *NoDiff* and *PronDiff* trials were similar, a more differentiated picture can be obtained if the time course of looking unfolding over the trial is taken into consideration.

Growth curve analysis (Mirman, 2014; Mirman, Dixon, & Magnuson, 2008) offers the possibility to analyse detailed eye tracking time course data using multilevel models. This involves fitting a polynomial curve to the shifting gaze pattern across the trial, or a certain time window within the trial. Compared to conducting statistical analyses for instance over a series of individual time-bins of looking data, the advantage is that the gaze pattern is captured by just a few parameters that define the curve, and the question of whether the curves for two types of trials differ statistically can be answered by the model. Growth curve analyses have recently become popular for eye tracking analyses, particularly in the context of visual world paradigms that investigate the time course of looking at a target over distractors in the context of speech processing (e. g. Chow, Aimola Davies, & Plunkett, 2017; Ito, Pickering, & Corley, 2018; Kukona, 2020).

Due to the nature of the task, subjects' gaze fell either on the match first or the non-match first. Scan patterns typically began at the top item, meaning approximately 50% of the trials were match-first, the remainder non-match-first. In order to compare whether looking patterns unfolded differently for trials where the first item encountered was the match, we plotted gaze-patterns as the probability of looking at the first-fixated item, separately for match-first and non-match-first trials (see Figs. 3 and 4). Trivially, these gaze patterns begin at 1. They then show a dip when (at least on some trials) subjects move on to the second item, and may then show a rise again as subjects return to the first item. The degree to which there is a dip and/or return to the first item is modulated by the trial type, as we shall see below.

For the growth curve analyses we needed to first decide on the specific models to be fitted, identify a suitable time window, and determine the order of the most suitable polynomial to be fitted to the gaze patterns. The aim of the modelling process is to determine the best-fitting curve, and this implies determining optimal shape parameters for a polynomial of a given order. Like any mixed-effects model, the process involves the selection of the "best model fit" on the basis of parameter optimisation. In addition to the fixed effects, so-called "time terms", the parameters for the polynomial, form a part of the fitting process as well. On the basis of visual inspection, we decided that 4th order polynomials would be appropriate in this particular case. In other words, the fitting

procedure involves four additional time parameters, the linear, quadratic, cubic and quartic time terms. Also based on visual inspection we selected the window of 1200 to 2200 ms after auditory onset, where 1200 ms corresponds approximately to the point at which subjects began to disengage from the first-fixated word.

Since match-first and non-match-first trials result in very different eye gaze patterns, it was not possible to fit growth curve models to both types of trials and all conditions simultaneously.[8] We therefore fitted two separate sets of models which complement each other:

(a) Discrimination models (one model per condition, comparing match-first vs. non-match-first gaze patterns)

An important indication of whether subjects are able to solve the task of matching the auditory fragment to the correct visual item is a difference in the time course of match-first vs. non-match-first trials in the same condition. If it cannot be decided whether a particular item is a match for the auditory fragment or not, the time course of trials where subjects started on either item (i.e. either match-first or non-match first) should look identical. As we will see below, that is exactly what we find for the *NoDiff* condition. By contrast, if subjects can easily decide whether they are looking at a match or not, then the two types of trials (match-first, non-match-first) differ, because a matching item, recognised as such, holds a subject's gaze for longer than a non-matching item, which can be rejected rapidly. We thus expect an easily decidable condition to result in a larger discrepancy between match-first and non-match-first trials, compared to a condition that is harder to decide. While manual accuracy scores give us information about whether, on average, subjects did or did not manage to identify the matching item, these curves can give us more fine-grained insight into how easy or difficult the discrimination task was. As a first instance we therefore establish whether the eye movement patterns differ significantly depending on whether the subjects' gaze fell first on the match or non-match. In total, there are therefore three Discrimination models, using fixed factors *Direction of first fixation* (match-first vs. non-match-first) and *Person* ("Discrimination models", see Fig. 3). We predicted that for the *NoDiff* condition, there would not be a difference between match- and non-match-first trials because it should not be possible to decide which of the written forms the auditory fragment corresponds to. By contrast, we predicted that the auditory fragments in the *PronDiff* and the *OrthPronDiff* conditions would provide enough cues to make a decision, i.e. in all other conditions match-first and non-match-first trials should have clearly distinct gaze patterns.

(b) Condition comparison models (one model for match-first trials across all conditions and one model for non-match-first trials across all conditions)

A comparison across conditions is necessary to quantify to what extent gaze patterns are different across conditions. Here we ask, for example, whether subjects' gaze was captured by the matching item systematically for a longer time in one condition or the other. To compare separately how the condition affected looking in match-first vs. non-match-first trials, we therefore fitted two separate models with fixed factors *Condition* and *Person*. We expected differences between the *NoDiff* condition and the remaining conditions in both trial types. We also predicted that looking patterns in the *OrthPronDiff* condition should exhibit the fastest non-match rejection in non-match-first trials and the most consistent looking at the match (i.e. lack of moving away) in the match-first trials. We expected looking patterns in the *PronDiff* condition to fall between the *NoDiff* and the *OrthPronDiff* condition in both match-first and non-match-first trials, reflecting the fact that while the auditory

---

[8] Attempts to include both match- and non-match-first-trials and all conditions in the same model led to poorly fitting models.
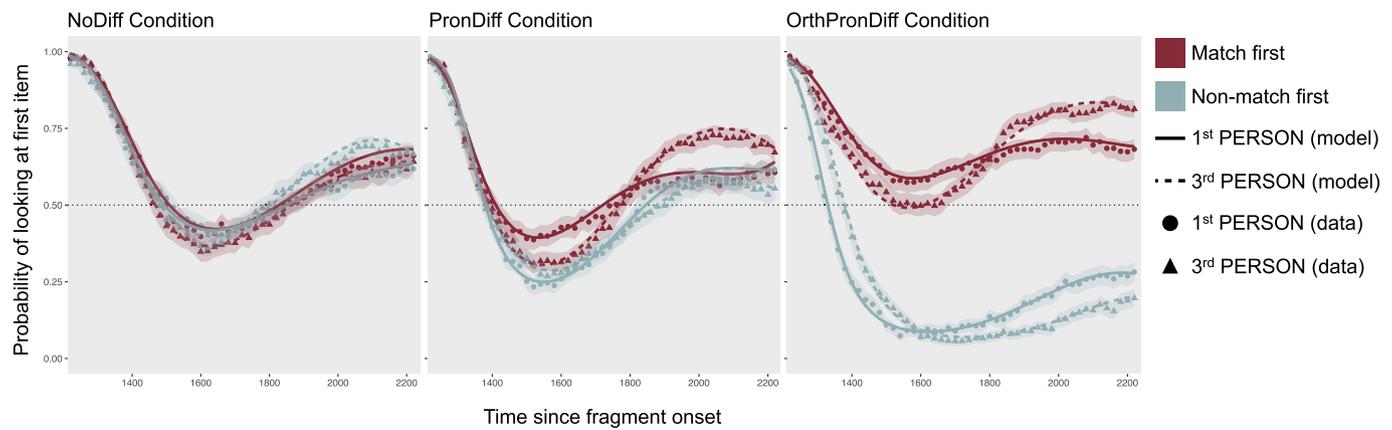
Fig. 3. Discrimination models: Data for *NoDiff* (a), *PronDiff* (b) and *OrthPronDiff* (c) conditions, with model predictions (separate models testing the effect of *Direction of first fixation and Person* in each condition). Red items represent Match-first trials, teal items represent Non-match-first trials. Lines depict model predictions (solid: 1st PERSON, dashed: 3rd PERSON), circles show data from 1st PERSON trials and triangles show data from 3rd PERSON trials. Shaded areas illustrate standard errors of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
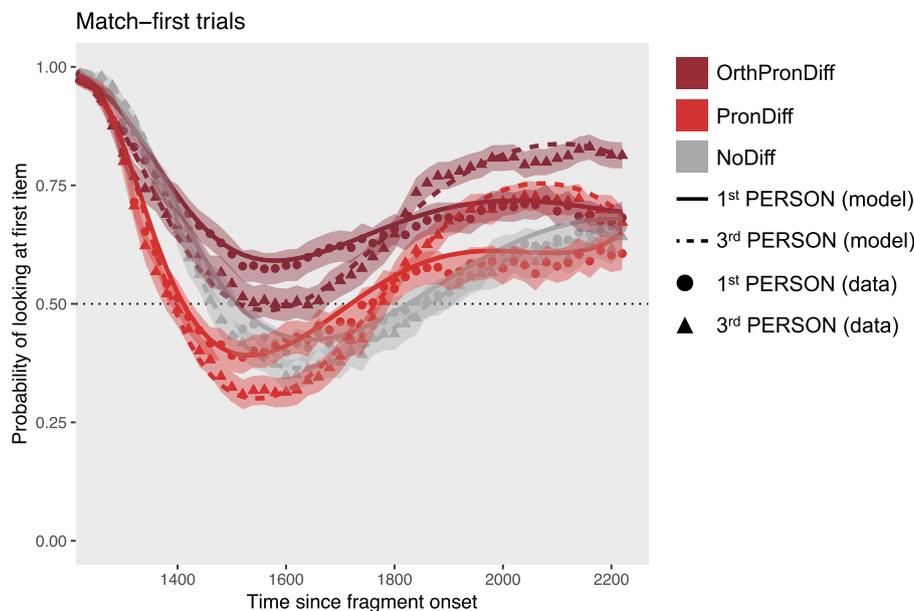

Fig. 4. Condition comparison model for match-first trials: Data and model predictions. Shaded areas show standard errors of the mean.

fragment allows a matching to morphological form, the visual representation generated on the basis of hearing the fragment matches both visually presented targets to some extent.

In order to conduct the growth curve analyses, we fitted logistic mixed effects models with the R-package lme4 (Bates et al., 2015) to the data. As described above, corresponding to the curved pattern in the data, we used fourth-order orthogonal polynomials, and modelled the time window between 1200 ms and 2200 ms after the onset of the auditory stimulus. We began with a base model containing only the time terms and crossed random effects of participants and items on intercepts. We then added fixed factors *Direction of first fixation* and *Person* as well as interactions[9] in a step-wise fashion, using model comparisons to

evaluate the improvements in model fit. Below we present only summaries of the statistical modelling results, with detailed results presented in Appendix B.

3.2.3.1. Models testing discrimination of match vs. non-match (Discrimination models). These models were fitted to establish subjects' ability to perform the task of fragment completion in each condition (separately). The rationale is that if it matters whether match or non-match are fixated first, with match-first fixations leading to a higher proportion of looking at that item, then subjects are able to discriminate between match and non-match. Final models included fixed effects of *Direction of first fixation* (match-first, non-match-first) and *Person* (1P, 3P) and interactions on all time terms, and random effects of subjects and items on the intercepts. Match-first trials were treated as the baseline, as were 1st PERSON trials. Detailed results of model comparisons and final model (parameter estimates) are provided in Appendix B.

In the *NoDiff* condition (Fig. 3a) neither the addition of an effect of *Direction of first fixation* ($\chi^2(1) = 1.3455$, $p > .24$) nor of an effect of *Person* ($\chi^2(1) = 0.053$, $p > .81$) on the intercept improved the model fit. This reflects the fact that both items were possible matches for the

fragment. The addition of an interaction of *Direction of first fixation* and *Person* improved the model ($\chi^2(1) = 38.96$, $p < .0001$). This reflects a higher likelihood of looking back at a 1st PERSON item if it was encountered first, consistent with the 3rd PERSON being considered a default. The addition of an effect of *Direction of first fixation* on the quadratic time term also significantly improved the model fit, as did the addition of an effect of *Person* on the linear and cubic terms, and the addition of an interaction on the linear and quartic terms (see Appendix B, Table 8). These changes to the shape of the curve also reflect the 3rd PERSON default status. See Appendix B, Table 9 for parameter estimates for the best-fitting model.

In the *PronDiff* condition (Fig. 3b), the effect of *Direction of first fixation* on the intercept significantly improved the model fit ($\chi^2(1) = 241.28$, $p < .0001$). Clearly, participants' eye movements in response to seeing either the match or the non-match first were distinct in the *PronDiff* condition, indicating that participants discriminated these items. The effect of *Person* only had an effect on some of the time terms, but not on the intercept ($\chi^2(1) = 0.0013$, $p > .97$). However, the addition of an interaction of *Direction of first fixation* and *Person* on the intercept yielded a significant improvement ($\chi^2(1) = 5.22$, $p = .022$), reflecting the fact that for match-first trials *Person* played a role, but not for non-match-first trials (see below). The addition of the effects on some of the time terms also improved the model fit, in particular the addition of an effect of *Direction of first fixation* on the quadratic term, the effect of *Person* on the linear and quartic terms and the addition of the interaction on the linear, quadratic and cubic time terms (for details see Appendix B, Table 10). The best-fitting model overall was therefore the one with these time terms on the interactions, see Appendix B, Table 11.

For *OrthPronDiff* trials (Fig. 3c), the effect of *Direction of first fixation* on the intercept significantly improved model fit ($\chi^2(1) = 8812.2$, $p < .0001$). The effect of *Person* on the intercept failed to improve model fit (intercept: $\chi^2(1) = 0.0009$, $p > .97$), but the addition of an interaction term of *Direction* and *Person* on the intercept yielded a trend ($\chi^2(1) = 2.83$, $p = .09$). These results imply that subjects were clearly able to distinguish between match and non-match in this condition. There were improvements in model fit also for the addition of effects on the time terms, in particular the addition of *Direction of first fixation* on the linear and quadratic time term, of *Person* on the quadratic, cubic and quartic time terms and the addition of the interaction on the linear, quadratic and cubic time terms, implying that the shape of the curve was also different depending on the two factors. Detailed results of the model comparison and parameter estimates for the final model are provided in Appendix B (Tables 12 & 13).

In summary, these three models demonstrate that subjects were able to distinguish match and non-match items not just for the *OrthPronDiff* condition, but also for the *PronDiff* condition (but not in the undecidable *NoDiff* condition). In other words, the eye movement patterns allow for more detailed insight into subjects' processing compared to the manual responses.

These separate models, however, do not allow us to compare the time course of looking across *OrthPronDiff* and *PronDiff* conditions, and thereby to determine whether they differ in terms of how easy matching the fragment to the visual target is. In order to do this, we turn to models that contain data from all three conditions, though separate for match-first and non-match-first trials.

3.2.3.2. Models comparing the time course of looking across conditions (Condition comparison models). As described above, in order to understand better to what extent the time course of looking differs across the conditions, we fitted two separate models to the match-first trials vs. the non-match-first trials. Final models included fixed effects of *Condition* (*NoDiff*, *PronDiff*, *OrthDiff*) and *Person* (1P, 3P) on all time terms, and random effects of subjects and items on the intercepts. The *NoDiff* condition was treated as a baseline in the models, as was 1st PERSON. Fixed effects were added in a step-wise fashion and model comparisons were

used to evaluate the improvements in model fit. Fig. 4 shows data and predictions for the best-fitting model for match-first trials, and Fig. 5 the corresponding for non-match-first trials. Results for model comparisons are summarised below and provided in detail in Appendix B, Table 14 (match-first) and Table 16 (non-match-first). Parameter estimates for best-fitting models are given in Tables 15 (match-first) and Table 17 (non-match-first).

Match-first trials The addition of an effect of *Condition* on the intercept improved the model fit ($\chi^2(2) = 57.72$, $p < .0001$). The addition of an effect of *Person* on the intercept did not improve the model fit ($\chi^2(1) = 0.539$, $p > .46$), and nor did the addition of a *Condition* x *Person* interaction ($\chi^2(2) = 2.205$, $p > .33$). The addition of the effect of *Condition* on the linear, quadratic, cubic and quartic time terms also improved the models (see Appendix B, Table 14), as did the addition of *Person* on all time terms. The addition of the interaction on linear and quartic time terms also yielded further improvements.

Inspecting the best-fitting model (see Appendix B, Table 15, for model estimates with respect to base *NoDiff*) reveals in particular that the main effect of *Condition* on the intercept shows a significant difference for the *OrthPronDiff* condition ($p < .0006$) compared to the base *NoDiff*, and a trend for the difference between *PronDiff* and *NoDiff* conditions. The difference between *OrthDiff* and *PronDiff* is also significant ($p < .0001$).[10]

Non-match-first trials The effect of *Condition* on the intercept improved the model fit ($\chi^2(2) = 206.26$, $p < .0001$). The addition of the effect of *Condition* on the linear, quadratic, cubic and quartic time terms led to further improvements (see Appendix B, Table 16). The addition of an effect of *Person* ($\chi^2(1) = 0.32$, $p > .57$) or of an interaction term ($\chi^2(2) = 1.271$, $p > .53$) on the intercept failed to do so, although adding the effect of *Person* on the linear, quadratic and quartic time terms did improve the fit and the addition of the interaction to the linear, quadratic and cubic time terms did as well.

In the best-fitting model (cf. Table 17), the difference between base *NoDiff* and *OrthPronDiff* conditions was significant ($p < .0001$), as was the difference between *NoDiff* and *PronDiff* ($p = .007$). *OrthPronDiff* and *PronDiff* conditions were also significantly different ($p < .0001$).

3.3. Discussion

As the models testing *Discrimination* demonstrate, subjects distinguish between match and non-match in both the *OrthPronDiff* and the *PronDiff* condition, despite the low manual response performance and overall looking time results in the *PronDiff* condition. The models testing the effect of *Condition* further confirm that *PronDiff* and *OrthPronDiff* conditions exhibit looking patterns that are different from each other. The *OrthPronDiff* condition seems easy to decide, with very distinct match-first vs. non-match-first curves, and high rates of looking at the match in both types of trial (remaining on the match when this is the first visual item encountered and rapidly rejecting the non-match when that is encountered first). The *PronDiff* condition seems more difficult to decide, with gaze patterns reflecting more pronounced shifts between the two visual items, and in particular very rapid movement away from the first item. Nevertheless, differences between match-first and non-match-first trials are clearly present, so subjects discriminate between the items, with an overall higher likelihood of remaining at or returning to the first item if this is the fully matching form. Considering the manual responses in this condition were at chance, this is quite remarkable. Eye movements reveal that subjects were in fact able to match items from this condition, contradictory to their low manual

---

[10] Lme4 provides estimates and *p*-values for contrasts against a specific base condition. We ran the models with the *NoDiff* condition as the base. Contrasts not involving this condition were obtained by using lme4's "relevel" function, which allows to output estimates and *p*-values for the same model but using a different base condition, in particular here the *PronDiff* condition.
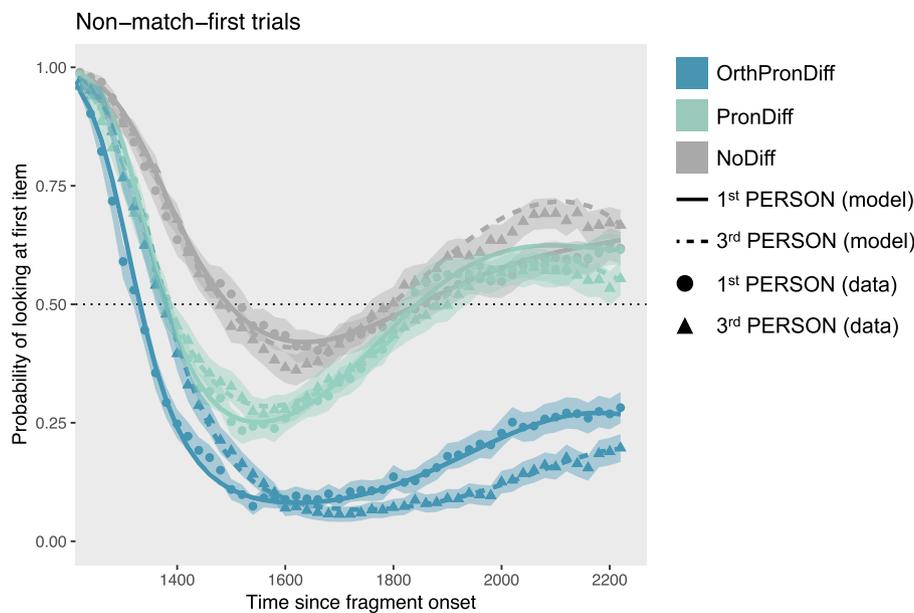
Fig. 5. Condition comparison model for non-match-first trials: Data and model predictions. Shaded areas show standard errors of the mean.

response accuracy.

Models also showed a difference between 1st PERSON and 3rd PERSON trials, as the *Condition* x *Person* interaction on time terms shows for match-first trials. Inspecting the plots in Fig. 4, it appears that the differences between 3rd PERSON and 1st PERSON are very similar across the *OrthPronDiff* and the *PronDiff* condition, which otherwise appear merely shifted in terms of the intercepts. There is a general tendency for subjects' gaze shifting towards the 3rd PERSON item, whether this is the target or not. This results in higher looking proportions at the match for 3rd PERSON trials towards the end of the trials, and lower looking proportions at the match for 1st PERSON trials. We believe that this is caused by the 3rd PERSON's status as the base form.

Inspecting the three conditions side by side shows that in the condition with an easy decision (*OrthPronDiff*) match-first and non-match-first trials result in the biggest difference, with long lingering on match and quick rejection of the non-match. The undecidable condition (*NoDiff*) results in shorter looking at the first item before moving away, but the hard-to-decide condition (*PronDiff*) shows even shorter looking at the first target – for both match- and non-match-first trials. In terms of the two alternative hierarchies given above, the data clearly correspond better to the second hypothesis (Hierarchy 2), where orthography plays a major role and the stem vowel ambiguity is less important. Trials in the *OrthPronDiff* condition were easiest to decide, as shown by the largest differences between match-first and non-match first trials (Fig. 3c). The precise time course in comparison to other conditions is most visible in match-first trials in Fig. 4, up until about 1800 ms after the fragment onset. Here subjects' lingering on the match at the beginning of the trial reflects the hierarchy as predicted, with subjects' proportion of looking at the first item (i.e. lack of looking away) highest for *OrthPronDiff* 1st PERSON trials, followed by *OrthPronDiff* 3rd PERSON trials. In other words, subjects in *OrthPronDiff* trials show a higher likelihood of maintaining looking at the matching item if this is seen first. These looking patterns are mirrored in the non-match-first trials (Fig. 5), where *OrthPronDiff* trials display the lowest likelihood of remaining on the first item – in other words, subjects rapidly reject the non-match and move their eyes towards the match, where they largely remain. Again, this is a slightly stronger effect in the 1st PERSON trials than in the 3rd PERSON trials.

PronDiff trials were harder to decide, as is clear from Fig. 3(b). Figs. 4 and 5 allow a comparison of the time course across these conditions. For match-first trials, the likelihood of remaining on the first item (the match) is clearly lower than the corresponding trials in the *OrthPronDiff* condition, and this is mirrored in the non-match-first trials, where compared to the *OrthPronDiff* condition subjects initially reject the non-matching item to a similar extent, but then are quite likely to return. We believe this in particular reflects the uncertainty in this condition. Finally, Fig. 3(a) shows that the *NoDiff* items were undecidable.

Had the stem vowel ambiguity played a more important role in this hierarchy than orthographic transparency, we would have expected 3rd PERSON trials in the *PronDiff* condition to be more easily decidable than either 3rd PERSON *OrthPronDiff* trials or 1st PERSON *PronDiff* trials (with [e], [o] being ambiguous as a predictor of Person). This was not the case, however. We therefore conclude that the looking patterns we identified here are clear indications of the association with distinct vs. identical graphemes being the driving factor underlying the difficulty of morphological discrimination.

However, in the hypothesized Hierarchy 2 we did not predict the systematic differences we find between 1st and 3rd PERSON trials. There are two potential underlying causes for differences between the 1st PERSON and 3rd PERSON forms. One is the ambiguity of the stem vowel in terms of either its ability to predict the suffix, or its association with a specific grapheme (even if effects are not as strong as predicted by Hierarchy 1 proposed above), a second one is the fact that the 3rd PERSON as base form is more easily accessed (activated as soon as a lexical entry is activated). However, the stem vowel ambiguity differs across *PronDiff* and *OrthPronDiff* condition (*OrthPronDiff*: 1st PERSON [i], [u] are uniquely associated with 1st PERSON/3rd PERSON [e], [o] are ambiguous; *PronDiff*: 1st PERSON [o], [e] are ambiguous, 3rd PERSON [æ], [ɔ] are unambiguous). In contrast, the Person-related shifts in the looking patterns appear very similar across the *PronDiff* and *OrthPronDiff* condition. We therefore believe that differences between 1P and 3P forms are caused by the 3rd PERSON item being the base form, rather than reflecting the level of ambiguity of the fragment vowel.

4. General discussion

In this paper, we used a set of inflectional stimuli in Bengali to investigate the interaction between phonology and orthography to determine to what extent each of the domains impacts processing and lexical access. Unlike previous research in this domain, the paradigm under investigation involves a completely predictable phonological alternation (governed by the high-vowel suffix [-i]) for which orthographic transparency varies systematically. In contrast to previously addressed questions such as work on the processing and storage of regular and irregular past tense forms in English, for instance (e.g. Crepaldi et al., 2010; Marcus et al., 1995; Marslen-Wilson & Tyler, 1998; McClelland & Patterson, 2002; Pastizzo & Feldman, 2002; Pinker, 1991), vowel raising in Bengali inflection is completely regular: The stem vowels [æ, e, ɔ, o] raise to [e, i, o, u] in the context of the 1st PERSON suffix [-i] but remain the same as in the citation form (verbal noun ending in [-a]) in the 3rd PERSON form ([-e]). Only the alternations [e ~ i] and [o ~ u] are reflected in the writing system while in the case of [æ ~ e] and [ɔ ~ o] the orthography remains the same despite raising of the vowel. We thus asked to what extent this alternation and its representation in the orthography result in differences in the access and discrimination mechanisms.

Both studies we presented here therefore investigated processing of three different groups of verbs, those exhibiting no differences in stem vowel between 1st PERSON and 3rd PERSON (*NoDiff*), those exhibiting vowel raising but without a change in orthography (*PronDiff*), and those exhibiting vowel raising *and* a corresponding change in orthography (*OrthPronDiff*). The first method employed was a cross-modal lexical decision task (Experiment 1) as it is the most frequently used method to investigate differences in lexical access. However, as the prime and target pairs are semantically identical the likelihood of strong priming effects across all three groups of verbs was high, and results indeed confirmed that priming was similar in all three types of words. We therefore also used eye-tracking with fragment completion in a forced-choice response paradigm (Experiment 2), which (a) uses a more sensitive task as subjects have to *discriminate* between two existing forms, and (b) allows for the use of detailed time course analysis of looking patterns. While the lexical decision task (Experiment 1) provided a global assessment of lexical access, this second study allowed us to probe whether the subtle differences in our stimuli generate differences in processing patterns rather than simply in the speed of full lexical access.

In Experiment 1, the results indicate that priming of the verbal noun is attenuated when the stem vowels in prime and target do not match (i. e. after a 1st PERSON prime, but not after a 3rd PERSON prime). One might argue that this could be due to a lesser degree of overlap: when the auditory prime is processed, its orthographic representation is activated, and the grapheme representing the stem vowel is identical to that of the verbal noun in all three word groups for 3rd PERSON primes, but only in two word groups for 1st PERSON primes. However, if this was the reason for the differences between 1st PERSON and 3rd PERSON primes, then there should also be an interaction with *Condition*, as only 1st PERSON primes in the *OrthPronDiff* condition activate a different grapheme. Since no difference between the *NoDiff, PronDiff* and *OrthPronDiff* conditions was found in Experiment 1, the greater degree of facilitation observed after 3rd PERSON primes is likely to be caused by the more basic nature of the 3rd PERSON in the inflectional paradigm (see Bybee, 2001; Lahiri, 1982). The lexical decision task thus did not provide any difference in lexical access between the conditions – but this is not altogether surprising, given the strong semantic relatedness of primes and targets.

Experiment 2 offered more fine-grained insight into the differences in processing across the three verb groups. In this task, hearing a fragment was expected to activate an orthographic representation which could then be matched to one of two visual targets. By forcing a discrimination decision between these semantically identical forms, which could be made with more or less certainty depending on the type of word presented, we were able to identify subtle differences in processing.

In particular, what stands out is the discrepancy between the manual response data and eye movement patterns. Button-press responses for the *PronDiff* condition indicated that subjects found the task hard, and in fact on average performance did not exceed chance. At the same time the eye movement patterns showed diverging results for trials during which fixations first fell on the match vs. on the non-match. Clearly subjects did reject a non-match to a greater extent than a matching target, and they moved on faster than they did in the *NoDiff* condition, where both items are potential matches. This implies that even for words where there is only a vowel alternation on its own, the alternation allows rapid prediction of the relevant suffix even if this is a smaller effect than for the *OrthPronDiff* condition.

Taken together, our experiments demonstrate that phonology and orthography both impact on lexical access in inflectional morphology. If orthographic differences made the discrimination of forms easy, but phonology had no impact at all, we would expect performance in the *PronDiff* condition to be equal to the *NoDiff* condition. If, on the other hand, phonological differences were the main driving factor, we would expect performance in *PronDiff* and *OrthPronDiff* conditions to be similar. What we found in the eye tracking task, by contrast, was that all three conditions are clearly distinct: phonological differences only just allow subjects to discriminate between match and non-match, and the low-accuracy manual responses indicate that there is a lot of uncertainty about the discrimination decision, but when orthography changes as well, the decision becomes easy. Perhaps more surprising than the fact that orthographic contrasts make the decision easy is the fact that their absence made the decision so difficult. Manual performance in the *PronDiff* condition was at chance, indicating that even though subjects' eye movements suggest differential processing of match and non-match in the immediate aftermath of hearing the fragment, the decision is a hard one to make. Clearly the phonological contrast is very subtle. Importantly though, in terms of phonological similarity there is no reason to suggest that the two forms in the *OrthPronDiff* condition are easier to discriminate – any gains in this condition in terms of performance are due to the fact that there are contrasting graphemes involved.

What is the implication for the mental activation patterns involved? After hearing the auditory fragment (but before processing the visual targets), are subjects equally certain in the *PronDiff* and the *OrthPronDiff* condition about whether they heard a part of a 1st PERSON or a 3rd PERSON item? Arguably, differences between the *PronDiff* and the *OrthPronDiff* condition could arise afterwards during the visual identification of the target item – i.e. it could be the visual matching process that is harder in the *PronDiff* condition (by definition there are fewer cues here than in the *OrthPronDiff* condition). However, if that were the only reason for a discrepancy between *PronDiff* and *OrthPronDiff* conditions, we would certainly expect above-chance manual performance in the *PronDiff* condition. To illustrate, we propose that the processes occurring upon presentation of the auditory fragment are as follows: (1) hearing the fragment activates a lexical entry (corresponding to the stem), (2) this retrieves an associated paradigm, i.e. a set of morphological features with phonological realisations. (3) This allows selecting the best match, i.e. identification of the precise morphological features, which also results in fragment completion, i.e. activation of the appropriate suffix and finally (4) a corresponding orthographic representation. (5) Once the visual stimuli are presented, the mental representation can be used to accept or reject the visual items. Note that the result of steps (1) and (2) is that both potential target items (i.e. 1st and 3rd PERSON forms) are

partially activated, but (3) narrows the set of competitors down to a single item where possible. It is this process that differs between the *PronDiff* and *OrthPronDiff* conditions.

To illustrate with an example, once the syllable [kʰe-] is heard, this activates the lexical entry for /kʰæla/ 'play'(1). The fragment would not be able to activate a lexical entry for */kʰela/because it does not exist in the lexicon. The activation of the lexical entry means that the set of morphological forms associated with this lexical entry becomes accessible, including the 1st and 3rd PERSON forms [kʰeli] and [kʰæle] (2). This constitutes partial activation of both forms, but [kʰeli] is then selected as best match, and this selection process constitutes the completion of the form, i.e. the prediction of the suffix [-i] (3). Now that one specific form is expected (both stem and suffix), the orthographic form খেলি is generated (4). Once a visual stimulus is encountered it is either matched (খেলি) or rejected (খেলে) (note the difference between the two Bengali grapheme sequences represents the suffix vowel, not the stem).

As already mentioned we believe our data show that the critical part in this sequence of processes, which differs across the *OrthPronDiff* and *PronDiff* conditions, is step (3), the decision about which exact morphological form has been heard (and therefore which suffix is required). The presentation of any 1st or 3rd PERSON fragment necessarily co-activates partially the other, due to phonological overlap (onset sounds) and subsequently semantic overlap once the lexical entry is activated. The difficulty of the decision about which form has been heard is therefore a function of the degree of overlap in the mental representation of 1st and 3rd PERSON. In terms of phonetics [kʰe-] is as different from [kʰæ-] as [li-] is from [le-] (from [likʰ-i]/[lekʰ-e], 1st and 3rd PERSON forms in the *OrthPronDiff* condition). The difference between the *OrthPronDiff* and *PronDiff* conditions is therefore only explained if the representation of the two vowels by different graphemes in the *OrthPronDiff* condition is useful. In other words, an advantage results from the mental representations of the stem vowels at this stage being more easily separable because they have in the past been encountered with two distinct graphemes. In other words, we argue here that the phonological representations are orthographically influenced.

There are no reasons to assume that (1) or (2) should be different across the two conditions, and once a full phonological form has been activated (i.e. fragment completion/morphological discrimination have occurred in step 3), generating an orthographic representation (4) is fully deterministic and ought to be equivalent across the conditions. If the differences only arose during visual matching (5), then the decision in the *PronDiff* condition might be harder than in the *OrthPronDiff* condition (due to the greater visual overlap between 1st and 3rd PERSON in the orthographic representation), but it would be *feasible* to make that decision as the suffix is available and unambiguous. It is not clear how chance-level manual responses would then be explained. As discussed above, there are no differences in phonological similarity between the pairs of stem vowels across the two conditions. We therefore argue that due to the existence of an orthographic contrast subjects are more certain about which form they heard in the *OrthPronDiff* condition.

Is it nevertheless possible that differences between *OrthPronDiff* and *PronDiff* conditions only arise here for task-based reasons, i.e. an orthographic representation is only activated because participants need it for visual matching with the targets? In other words, is it possible that steps (4)/(5) are really the source of the critical differences, but in a naturalistic listening scenario these would not be required for processing, and step (3) is in fact not the locus of the difficulty at all? We believe our data hold the answer to this. Let us assume the decision process is solely different across the conditions because of an activated orthographic representation. Then the *OrthPronDiff* condition should be easier to decide than the *PronDiff* condition (because the representation of two distinct graphemes facilitates the process). But in the *PronDiff* condition,

words with a front and back vowel should present an equal level of difficulty because in both cases the stem vowel is represented identically across the 1st and 3rd PERSON (i.e. the stem vowel cannot help with the decision). We would therefore expect no differences between *PronDiff* words with front vs. back vowels. Yet that is not what we found: subjects displayed above-chance accuracy for words with back vowels, which are *not at all* represented graphemically, but their responses did not exceed chance for words with front vowels, which are represented by the *same* grapheme. In other words, overlapping stem vowel representations interfere more than if the stem vowel is not represented. This, in our view, is evidence that the locus of the difficulty is the (orthographically influenced) phonological representation, and not an orthographic representation per se: Back vowel words in the *PronDiff* condition are not affected, either positively or negatively, by orthography because the stem vowels are not written. Front vowels are written with the same grapheme, and this makes phonological mental represenatations such as those of [kʰæl-e] and [kʰel-i] highly confusable. *OrthPronDiff* words are written with two distinct graphemes and this makes the phonological representation of /lekʰ-e/ and /likʰ-i/ highly distinct.

Of course, while the vowel type differences imply that orthographically influenced phonological representations are the main underlying factor determining discrimination in this task, we cannot entirely exclude the possibility that task-induced orthographic activation contributes to the effect, or its strength. In order to do this, one would have to conduct an experiment that does not rely on visual stimuli – a goal for future research.

However, as explained above, we conclude that the critical step is fragment completion, and this resolution of the morphological form has to occur in speech processing independent of whether there is a grapheme-based task. Since Bengali is a verb-final PRO-drop language there is no argument to be made that in spoken language the relevant information may not have to be decoded from the morphological form – on the contrary, verb morphology has a crucial role.

Our interpretation is that the very existence of an orthographic contrast has over time sensitised listeners to the phonological contrast between the alternating forms, meaning that subjects are more likely to notice the vowel alternation during processing. In other words, encountering two distinct graphemic realisations over and over again in the context of a vowel alternation (in the *OrthPronDiff* condition, but not in the *PronDiff* condition) may have shaped the phonological representation itself. Alternatively the automatic co-activation of an orthographic representation means that overall activation patterns in the *OrthPronDiff* condition are more disctict. Our study cannot discriminate between these two possibilities. What is certain is that orthographic alternations have a clear impact on phonological discrimination in the context of inflectional morphology. While Bengali vowel raising provides an ideal testing scenario for such an interaction, due to its high level of regularity, this has implications for morphophonological processes in general. It is difficult to capture such effects in languages such as English or German experimentally because of the prevalence of vowel alternations in irregular forms rather than across large categories of words and because the simultaneous effects of orthography and phonology are less frequent and less regular. Our results imply that phonological alternations can be used by listeners to predict upcoming suffix information, allowing the morphological form to be resolved more quickly – in other words, they contribute to efficient processing. However, if solely phonological in nature, these benefits are rather subtle and whether subjects make use of them in natural speech processing seems doubtful – it is orthographic transparency that provides the real boost. Spelling alternations clearly have a profound impact on the discrimination of morphophonological forms.

Author credit statement

Nadja Althaus: Conceptualization, Data curation, Methodology, Formal analysis, Investigation, Visualisation, Writing – original draft, Writing, reviewing & editing

Sandra Kotzor: Conceptualization, Data curation, Methodology, Formal analysis, Investigation, Visualisation, Writing – original draft, Writing, reviewing & editing

Swetlana Schuster: Methodology, Data curation

Aditi Lahiri: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing, reviewing & editing

Appendix A. Bengali transcription and full stimulus list

Table 5

PronDiff word list (low-to-mid vowel change).

	3P.SG	1P.SG	Target Verbal noun	Meaning	Bengali Target
1.	kʰæl-e	kʰel-i	kʰæl-a	play	খেলা
2.	ɖækʰ-e	ɖekʰ-i	ɖækʰ-a	see	দেখা
3.	bæl-e	bel-i	bæl-a	roll pastry	বেলা
4.	pʰæl-e	pʰel-i	pʰæl-a	throw	ফেলা
5.	tʰæl-e	tʰel-i	tʰæl-a	push	ঠেলা
6.	bætʃ-e	betʃ-i	bætʃ-a	sell	বেচা
7.	mæl-e	mel-i	mæl-a	hang up laundry	মেলা
8.	ʃæ̃k-e	ʃẽk-i	ʃæ̃k-a	dry fry	সেঁকা
9.	tʃʰæ̃k-e	tʃʰẽk-i	tʃʰæ̃k-a	fry	ছেঁকা
10.	tʰæk-e	tʰek-i	tʰæk-a	touch	ঠেকা
11.	tʰæʃ-e	tʰeʃ-i	tʰæʃ-a	push against	ঠেসা
12.	nɔɽ-e	noɽ-i	nɔɽ-a	totter, move	নড়া
13.	ʃɔr-e	ʃor-i	ʃɔr-a	move away	সরা
14.	tʃɔɽ-e	tʃoɽ-i	tʃɔɽ-a	climb	চড়া
15.	bʰɔɽ-e	bʰoɽ-i	bʰɔɽ-a	fill	ভড়া
16.	bɔk-e	bok-i	bɔk-a	scold	বকা
17.	tʃɔl-e	tʃol-i	tʃɔl-a	walk	চলা
18.	kɔr-e	kor-i	kɔr-a	do	করা
19.	dʰor-e	dʰor-i	dʰɔr-a	hold	ধরা
20.	gɔɽ-e	goɽ-i	gɔɽ-a	start	গড়া
21.	bɔl-e	bol-i	bɔl-a	say	বলা
22.	pɔɽ-e	poɽ-i	pɔɽ-a	read	পড়া
23.	bɔʃ-e	boʃ-i	bɔʃ-a	sit	বসা
24.	bæ̃k-e	bẽk-i	bæ̃k-a	bend	বেঁকা

Table 6
OrthPronDiff word list (mid-to-high vowel change).

	3P.SG	1P.SG	Target Verbal noun	Meaning	Bengali Target
1.	kʰɔ̃ɽ-e	kʰũɽ-i	kʰɔ̃ɽ-a	dig	খোঁড়া
2.	d̪ʰok-e	d̪ʰuk-i	d̪ʰok-a	enter	ঢোকা
3.	ʃon-e	ʃun-i	ʃon-a	hear	শোনা
4.	kʰol-e	kʰul-i	kʰol-a	open	খোলা
5.	põtʃʰ-e	pũtʃʰ-i	põtʃʰ-a	wipe	পোঁছা
6.	ʃõk-e	ʃũk-i	ʃõk-a	smell	শোঁকা
7.	ɖol-e	ɖul-i	ɖol-a	swing	দোলা
8.	motʃʰ-e	mutʃʰ-i	motʃʰ-a	wipe	মোছা
9.	kʰõdʒ-e	kʰũdʒ-i	kʰõdʒ-a	search, find	খোঁজা
10.	bodʒʰ-e	budʒʰ-i	bodʒʰ-a	understand	বোঝা
11.	dʒeʈ-e	dʒiʈ-i	dʒeʈ-a	win	জেতা
12.	dʒʰol-e	dʒʰul-i	dʒʰol-a	hang down	ঝোলা
13.	ʈol-e	ʈul-i	ʈol-a	hold up	তোলা
14.	tʰok-e	tʰuk-i	tʰok-a	knock	ঠোকা
15.	ɖob-e	ɖub-i	ɖob-a	dive	ডোবা
16.	tʃoʃ-e	tʃuʃ-i	tʃoʃ-a	suck	চোষা
17.	ʈep-e	ʈip-i	ʈep-a	press	টেপা
18.	ʃekʰ-e	ʃikʰ-i	ʃekʰ-a	learn	শেখা
19.	meʃ-e	miʃ-i	meʃ-a	mix	মেশা
20.	peʈ-e	piʈ-i	peʈ-a	beat up	পেটা
21.	tʃʰẽɽ-e	tʃʰĩɽ-i	tʃʰẽɽ-a	tear	ছেঁড়া
22.	tʃen-e	tʃin-i	tʃen-a	know someone	চেনা
23.	lekʰ-e	likʰ-i	lekʰ-a	write	লেখা
24.	pʰer-e	pʰir-i	pʰer-a	return	ফেরা

Table 7
NoDiff word list (low vowel /a/).

	3P.SG	1P.SG	Target Verbal noun	Meaning	Bengali Target
1.	kʰaʈ-e	kʰaʈ-i	kʰaʈ-a	work hard	খাটা
2.	mar-e	mar-i	mar-a	hit	মারা
3.	baʧʰ-e	baʧʰ-i	baʧʰ-a	separating grains	বাছা
4.	kaʈ-e	kaʈ-i	kaʈ-a	cut	কাটা
5.	rakʰ-e	rakʰ-i	rakʰ-a	keep	রাখা
6.	naʧ-e	naʧ-i	naʧ-a	dance	নাচা
7.	an-e	an-i	an-a	bring	আনা
8.	ʧap-e	ʧap-i	ʧap-a	press	চাপা
9.	hãʈ-e	hãʈ-i	hãʈ-a	walk	হাঁটা
10.	hãʃ-e	hãʃ-i	hãʃ-a	laugh	হাঁসা
11.	aʃ-e	aʃ-i	aʃ-a	come	আসা
12.	baʈ-e	baʈ-i	baʈ-a	grind	বাটা
13.	bʰaŋ-e	bʰaŋ-i	bʰaŋ-a	break	ভাঙ্গা
14.	bʰadʒ-e	bʰadʒ-i	bʰadʒ-a	fry	ভাজা
15.	ʧʰãk-e	ʧʰãk-i	ʧʰãk-a	strain, sieve	ছাঁকা
16.	bʰab-e	bʰab-i	bʰab-a	think	ভাবা
17.	ʧʰap-e	ʧʰap-i	ʧʰap-a	printing	ছাপা
18.	bãd̪ʰ-e	bãd̪ʰ-i	bãd̪ʰ-a	tie up	বাঁধা
19.	kãɖ-e	kãɖ-i	kãɖ-a	cry	কাঁদা
20.	ɖak-e	ɖak-i	ɖak-a	call out	ডাকা
21.	t̪ʰak-e	t̪ʰak-i	t̪ʰak-a	stay	থাকা
22.	ʈan-e	ʈan-i	ʈan-a	pull	টানা
23.	gãt̪ʰ-e	gãt̪ʰ-i	gãt̪ʰ-a	string pearls	গাঁথা
24.	ʃadʒ-e	ʃadʒ-i	ʃadʒ-a	dress up	সাজা

Appendix B. Models – full results

B.1. Discrimination models

The base model always contained only the four time terms and random effects of subjects and items on the intercept.

Table 8
NoDiff model comparisons.

Added effect	χ^2	df	p
Direction of First Fixation (Intercept)	1.35	1	0.246
Direction of First Fixation (Linear)	3.81	1	0.051
Direction of First Fixation (Quadratic)	7.92	1	0.005
Direction of First Fixation (Cubic)	0.45	1	0.502
Direction of First Fixation (Quartic)	1.03	1	0.309
Person (Intercept)	0.05	1	0.818
Person (Linear)	16.89	1	<0.0001
Person (Quadratic)	1.13	1	0.287
Person (Cubic)	11.66	1	0.0006
Person (Quartic)	0.65	1	0.421
Direction of First Fixation x Person (Intercept)	38.96	1	<0.0001
Direction of First Fixation x Person (Linear)	19.77	1	<0.0001
Direction of First Fixation x Person (Quadratic)	0.32	1	0.570
Direction of First Fixation x Person (Cubic)	0.09	1	0.762
Direction of First Fixation x Person (Quartic)	7.75	1	0.005

Table 9
Parameter estimates for best-fitting NoDiff model (w.r.t. base: Nonmatch-first/1P).

	Est.	SE	z	p
Intercept	0.52	0.13	4.10	<0.0001
Linear	−3.64	0.22	−16.77	<0.0001
Quadratic	5.63	0.22	25.21	<0.0001
Cubic	−2.98	0.21	−14.39	<0.0001
Quartic	0.67	0.19	3.56	0.0004
Direction of First Fixation (MF) Intercept	0.11	0.04	2.97	0.003
Direction of First Fixation (MF) Linear	0.39	0.31	1.26	0.209
Direction of First Fixation (MF) Quadratic	0.31	0.31	0.98	0.329
Direction of First Fixation (MF) Cubic	0.06	0.29	0.21	0.838
Direction of First Fixation (MF) Quartic	−0.32	0.27	−1.19	0.23
Person (3P) Intercept	0.09	0.07	1.34	0.181
Person (3P) Linear	1.46	0.30	4.86	0.000
Person (3P) Quadratic	0.07	0.31	0.23	0.815
Person (3P) Cubic	−0.33	0.29	−1.15	0.249
Person (3P) Quartic	−0.65	0.26	−2.49	0.013
Direction of First Fixation x Person Intercept	−0.19	0.05	−3.73	0.000
Direction of First Fixation x Person Linear	−2.14	0.44	−4.86	0.000
Direction of First Fixation x Person Quadratic	0.80	0.45	1.77	0.077
Direction of First Fixation x Person Cubic	−0.60	0.42	−1.44	0.149
Direction of First Fixation x Person Quartic	1.06	0.38	2.80	0.005

MF = Match-first.

Table 10
PronDiff model comparisons.

Added effect	χ^2	df	p
Direction of First Fixation (Intercept)	241.28	1	<0.0001
Direction of First Fixation (Linear)	3.51	1	0.061
Direction of First Fixation (Quadratic)	14.37	1	0.0001
Direction of First Fixation (Cubic)	1.09	1	0.297
Direction of First Fixation (Quartic)	1.06	1	0.302
Person (Intercept)	0.001	1	0.971
Person (Linear)	24.58	1	<0.0001
Person (Quadratic)	1.34	1	0.247
Person (Cubic)	2.89	1	0.089
Person (Quartic)	13.82	1	0.0002
Direction of First Fixation x Person (Intercept)	5.22	1	0.022
Direction of First Fixation x Person (Linear)	73.17	1	<0.0001
Direction of First Fixation x Person (Quadratic)	21.41	1	<0.0001
Direction of First Fixation x Person (Cubic)	44.75	1	<0.0001
Direction of First Fixation x Person (Quartic)	0.10	1	0.748

Table 11
Parameter estimates for best-fitting PronDiff model (w. r. t. base: Non-match-first/1P).

	Est.	SE	z	p
Intercept	0.16	0.11	1.45	0.148
Linear	−1.41	0.19	−7.39	<0.0001
Quadratic	6.11	0.20	30.39	<0.0001
Cubic	−4.97	0.19	−25.53	<0.0001
Quartic	1.38	0.16	8.77	<0.0001
Direction of First Fixation (MF) Intercept	0.25	0.03	7.35	<0.0001
Direction of First Fixation (MF) Linear	−0.34	0.26	−1.31	0.191
Direction of First Fixation (MF) Quadratic	−1.91	0.27	−6.97	<0.0001
Direction of First Fixation n(MF) Cubic	1.35	0.26	5.117	<0.0001
Direction of First Fixation (MF) Quartic	0.15	0.18	0.83	0.406
Person (3P) Intercept	−0.08	0.09	−0.95	0.342
Person (3P) Linear	0.07	0.26	0.29	0.771
Person (3P) Quadratic	−1.16	0.27	−4.34	<0.0001
Person (3P) Cubic	1.15	0.26	4.40	<0.0001
Person (3P) Quartic	−0.69	0.18	−3.869	0.0001
Direction of First Fixation x Person Intercept	0.15	0.05	3.19	0.0014
Direction of First Fixation x Person Linear	1.71	0.36	4.77	<0.0001
Direction of First Fixation x Person Quadratic	2.43	0.37	6.59	<0.0001
Direction of First Fixation x Person Cubic	−2.40	0.36	−6.69	<0.0001

MF = match-first.

Table 12
OrthPronDiff model comparisons.

Added effect	χ^2	df	p
Direction of First Fixation (Intercept)	8812.24	1	<0.0001
Direction of First Fixation (Linear)	296.51	1	<0.0001
Direction of First Fixation (Quadratic)	283.99	1	<0.0001
Direction of First Fixation (Cubic)	0.12	1	0.730
Direction of First Fixation (Quartic)	3.39	1	0.066
Person (Intercept)	0.0009	1	0.976
Person (Linear)	1.25	1	0.263
Person (Quadratic)	53.79	1	<0.0001
Person (Cubic)	4.56	1	0.033
Person (Quartic)	4.69	1	0.030
Direction of First Fixation x Person (Intercept)	2.83	1	0.093
Direction of First Fixation x Person (Linear)	353.64	1	<0.0001
Direction of First Fixation x Person (Quadratic)	4.2	1	0.040
Direction of First Fixation x Person (Cubic)	17.22	1	<0.0001
Direction of First Fixation x Person (Quartic)	0.007	1	0.933

Table 13
Parameter estimates for OrthPronDiff model (w.r.t. base: Non-match-first/1P).

	Est.	SE	z	p
Intercept	−1.23	0.10	−11.97	<0.0001
Linear	−2.90	0.18	−15.89	<0.0001
Quadratic	6.76	0.21	32.84	<0.0001
Cubic	−3.68	0.20	−18.08	<0.0001
Quartic	0.50	0.17	2.86	0.004
Direction of First Fixation (TF) Intercept	2.20	0.04	57.64	<0.0001
Direction of First Fixation (TF) Linear	0.56	0.27	2.05	0.040
Direction of First Fixation (TF) Quadratic	−3.17	0.29	−11.01	<0.0001
Direction of First Fixation (TF) Cubic	0.84	0.28	3.03	0.002
Direction of First Fixation (TF) Quartic	0.27	0.20	1.36	0.173
Person (3P) Intercept	−0.08	0.10	−0.83	0.408
Person (3P) Linear	−3.94	0.28	−14.31	<0.0001
Person (3P) Quadratic	1.38	0.31	4.42	<0.0001
Person (3P) Cubic	1.17	0.30	3.85	0.00012
Person (3P) Quartic	−0.55	0.20	−2.77	0.006
Direction of First Fixation x Person Intercept	0.09	0.06	1.69	0.091
Direction of First Fixation x Person Linear	6.83	0.39	17.67	<0.0001
Direction of First Fixation x Person Quadratic	−0.47	0.42	−1.12	0.262
Direction of First Fixation x Person Cubic	−1.68	0.40	−4.17	<0.0001

MF = Match-first.

B.2. Condition comparison models

Table 14
Match-first model comparisons.

Added effect	χ^2	df	p
Condition (Intercept)	57.72	2	<0.0001
Condition (Linear)	250.1	2	<0.0001
Condition (Quadratic)	162.65	2	<0.0001
Condition (Cubic)	32.1	2	<0.0001
Condition (Quartic)	17.96	2	0.0001
Person (Intercept)	0.54	1	0.463
Person (Linear)	163.12	1	<0.0001
Person (Quadratic)	22.03	1	<0.0001
Person (Cubic)	41.94	1	<0.0001
Person (Quartic)	4.81	1	0.028
Condition x Person (Intercept)	2.20	2	0.332
Condition x Person (Linear)	100.43	2	<0.0001
Condition x Person (Quadratic)	2.48	2	0.289
Condition x Person (Cubic)	5.47	2	0.065
Condition x Person (Quartic)	10.6	2	0.005

Table 15
Parameter estimates for best-fitting match-first model (w.r.t. base: NoDiff condition/1P).

	Est.	SE	z	p
Intercept	0.62	0.13	4.88	<0.0001
Linear	−3.26	0.21	−15.30	<0.0001
Quadratic	5.88	0.22	26.99	<0.0001
Cubic	−2.90	0.20	−14.4	<0.0001
Quartic	0.38	0.19	2.05	0.04
OrthPronDiff Intercept	0.37	0.11	3.41	0.0006
OrthPronDiff Linear	0.85	0.29	2.89	0.003
OrthPronDiff Quadratic	−2.23	0.30	−7.46	<0.0001
OrthPronDiff Cubic	0.06	0.28	0.22	0.82
OrthPronDiff Quartic	0.38	0.26	1.45	0.15
PronDiff Intercept	−0.19	0.11	−1.73	0.08
PronDiff Linear	1.49	0.29	5.22	<0.0001
PronDiff Quadratic	−1.53	0.29	−5.22	<0.0001
PronDiff Cubic	−0.83	0.28	−3.00	0.003
PronDiff Quartic	1.22	0.26	4.73	<0.0001
Person(3P) Intercept	−0.10	0.11	−0.90	0.37
3P Linear	−0.72	0.32	−2.26	0.02
3P Quadratic	0.97	0.33	2.95	0.003
3P Cubic	−1.01	0.30	−3.34	0.0008
3P Quartic	0.40	0.27	1.47	0.14
OrthPronDiff:3P Intercept	0.10	0.16	0.64	0.52
OrthPronDiff:3P Linear	0.16	0.15	1.01	0.31
OrthPronDiff:3P Quadratic	3.67	0.43	8.54	<0.0001
OrthPronDiff:3P Cubic	2.55	0.42	6.06	<0.0001
OrthPronDiff:3P Quartic	0.00	0.44	0.01	0.99
PronDiff:3P Intercept	0.28	0.43	0.65	0.51
PronDiff:3P Linear	0.40	0.41	0.99	0.32
PronDiff:3P Quadratic	−0.24	0.40	−0.58	0.56
PronDiff:3P Cubic	−0.91	0.38	−2.40	0.016
PronDiff:3P Quartic	−1.19	0.37	−3.18	0.001

Table 16
Non-match-first model comparisons.

Added effect	χ^2	df	p
Condition (Intercept)	206.26	2	<0.0001
Condition (Linear)	436.81	2	<0.0001
Condition (Quadratic)	161.46	2	<0.0001
Condition (Cubic)	42.24	2	<0.0001
Condition (Quartic)	22.03	2	<0.0001
Person (Intercept)	0.32	1	0.574
Person (Linear)	31.67	1	<0.0001
Person (Quadratic)	6.58	1	0.010
Person (Cubic)	0.1	1	0.757
Person (Quartic)	9.33	1	0.002
Condition x Person (Intercept)	1.27	2	0.530
Condition x Person (Linear)	230.82	2	<0.0001
Condition x Person (Quadratic)	34.11	2	<0.0001
Condition x Person (Cubic)	20.21	2	<0.0001
Condition x Person (Quartic)	0.039	2	0.981

Table 17
Parameter estimates for non-match-first model (w.r.t. base: NoDiff condition/1P).

	Est.	SE	z	p
Intercept	0.52	0.13	4.11	<0.0001
Linear	−3.59	0.21	−17.21	<0.0001
Quadratic	5.57	0.21	26.06	<0.0001
Cubic	−2.96	0.20	−14.94	<0.0001
Quartic	0.63	0.15	4.12	<0.0001
Condition/OrthPronDiff Intercept	−1.79	0.13	−13.67	<0.0001
Condition/OrthPronDiff Linear	0.52	0.28	1.88	0.06
Condition/OrthPronDiff Quadratic	1.45	0.30	4.83	<0.0001
Condition/OrthPronDiff Cubic	−0.87	0.29	−3.06	0.002
Condition/OrthPronDiff Quartic	−0.09	0.20	−0.45	0.66
Condition/PronDiff Intercept	−0.35	0.13	−2.70	0.007
Condition/PronDiff Linear	2.21	0.28	7.96	<0.0001

Table 17 (*continued*)

	Est.	SE	z	p
Condition/PronDiff Quadratic	0.52	0.29	1.83	0.07
Condition/PronDiff Cubic	−1.99	0.27	−7.35	<0.0001
Condition/PronDiff Quartic	0.69	0.18	3.81	0.0001
Person(3P) Intercept	0.11	0.13	0.88	0.38
3P Linear	1.43	0.28	5.10	<0.0001
3P Quadratic	0.04	0.29	0.15	0.88
3P Cubic	−0.33	0.27	−1.22	0.22
3P Quartic	−0.61	0.16	−3.93	<0.0001
OrthPronDiff:3P Intercept	−0.18	0.19	−0.96	0.34
OrthPronDiff:3P Linear	−5.43	0.39	−13.91	0.004
OrthPronDiff:3P Quadratic	1.24	0.43	2.92	<0.0001
OrthPronDiff:3P Cubic	1.61	0.40	4.00	<0.0001
PronDiff:3P Intercept	−0.2	0.18	−1.09	0.28
PronDiff:3P Linear	−1.4	0.37	−3.79	0.0002
PronDiff:3P Quadratic	−1.11	0.39	−2.93	0.003
PronDiff:3P Cubic	1.37	0.36	3.79	0.0002

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology, 3*, 1–12.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *The Quarterly Journal of Experimental Psychology, 42*(4), 693–711. https://doi.org/10.1080/14640749008401245

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological), 26*(2), 211–243.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Chang, E., Rieger, J., Johnson, K., Berger, M., Barbaro, N., & Knight, R. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*, 1428–1432.

Chatterji, S. K. (1926/1978). *The origin and development of the Bengali language*. Calcutta, New Delhi: Rupa & Co.. Original published 1926, Calcutta University.

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*(3), 341–360.

Chow, J., Aimola Davies, A., & Plunkett, K. (2017). Spoken-word recognition in 2-year-olds: The tug of war between phonological and semantic activation. *Journal of Memory and Language, 93*(October), 104–134.

Clahsen, H., Eisenbeiss, S., & Sonnenstuhl-Henning, I. (1997). Morphological structure and the processing of inflected words. *Theoretical Linguistics, 23*(3), 201–249.

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). "Fell" primes "fall", but does "bell" prime "ball"? Masked priming with irregularly-inflected primes. *Journal of Memory and Language, 63*(1), 83–99.

Cutler, A., & Davis, C. (2012). An orthographic effect in phoneme processing, and its limitations. *Frontiers in Psychology, 3*, 18.

Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 32*(3), 535–557.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., … Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*(6009), 1359–1364.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Journal of Memory and Cognition, 13*, 241–255.

Frost, R., & Ziegler, J. C. (2007). Speech and spelling interaction: The interdependence of visual and auditory word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 107–118).

Hock, H. H. (1991). *Principles of historical linguistics*. Berlin: De Gruyter Mouton.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11.

Kielar, A., Joanisse, M. F., & Hare, M. L. (2008). Priming English past tense verbs: Rules or statistics? *Journal of Memory and Language, 58*(2), 327–346.

Kiparsky, P. (1982). From cyclic to lexical phonology. In H. van der Hulst (Ed.), *The structure of phonological representation*. Dordrecht, The Netherlands: Foris.

Kiparsky, P., & Tonhauser, J. (2013). Semantics of inflection. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *3. Semantics. An international handbook of language meaning* (pp. 2070–2097). Berlin: De Gruyter Mouton.

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 46*(11), 2153–2162.

Lahiri, A. (1982). *Theoretical implications of analogical change: Evidence from Germanic languages*. Brown University. University Microfilms.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*, 245–294.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics, 38*, 44–59.

Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex, 116*, 4–44.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*, 189–256.

Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2*, 428e435.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*(1), 3–33.

McClelland, J., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*(11), 465–472.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Journal of Cognitive Psychology, 18*, 1–86.

McQueen, J. M., & Sereno, J. (2005). Cleaving automatic processes from strategic biases in phonological priming. *Memory & Cognition, 33*(7), 1185–1209.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science (New York, N.Y.), 343*(6174), 1006–1010.

Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language & Cognitive Processes, 19*(4), 561–580.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494.

Pastizzo, M. J., & Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*(1), 244–249.

Pattamadilok, C., Kolinsky, R., Ventura, P., Radeau, M., & Morais, J. (2007). Orthographic representations in spoken word priming: No early automatic activation. *Language and Speech, 50*(4), 505–531.

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language & Cognitive Processes, 22*(5), 700–726.

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research, 1275*, 73–80.

Pinker, S. (1991). Rules of language. *Science, 253*, 254–530.

Plank, F., & Lahiri, A. (2015). Macroscopic and microscopic typology: Basic valence orientation, more pertinacious than meets the naked eye. *Linguistic Typology, 19*(1), 1–54.

Reetz, H., & Kleinmann, A. (2003). Multi-subject hardware for experiment control and precise recaction time measurement. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences, Barcelona* (pp. 1489–1492).

Scharinger, M., Lahiri, A., & Eulitz, C. (2010). Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics, 23*(4), 383–399.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 546–554.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behaviour, 18*, 399–412.

Taft, M. (1991). *Reading and the mental lexicon*. Hove: Erlbaum.

Taft, M. (2011). Orthographic influences when processing spoken pseudowords: Theoretical implications. *Frontiers in Psychology, 2*(149), 1–7.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*(2), 366–379.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*(6), 513–520.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language & Cognitive Processes, 19*(1), 57–95.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683–689.

Ziegler, J. C., Petrova, A., & Ferrand, L. (2008). Feedback consistency effects in visual and auditory word recognition: Where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 643.