The importance of prediction in language processing

Falk Huettig

falk.huettig@mpi.nl
Prediction – a unifying principle of the mind?

- it “offers a deeply unified account of perception, cognition, and action”

- “brains … are essentially prediction machines” (A. Clark, 2013)

- an old idea (von Helmholtz, 1868; James, 1890)
Prediction – a unifying principle of the mind?

- perception & action research (Wolpert et al., 2003; Sebanz & Knoblich, 2009; von Hofsten, 2004; etc.)

- 6-month-olds anticipate (Hunnuis & Bekkering, 2010)

- predictive coding: brain is fundamentally engaged in predictive coding and computes precise prediction errors which bias towards making correct inferences (e.g., Friston, 2010)
(Some) linguists don’t think so

- “what good would such predictions do in understanding the sentences?”

- “predicting the next word has no bearing whatsoever on an explanation of speech production, where the goal has to be to produce the next word in an effort to say something meaningful”

(Jackendoff, 2007)
4 central questions

Why?
• what is the function of prediction?

What?
• cues used for prediction
• contents of prediction

How?
• mechanisms of prediction
• mediating factors

When?
• do people always predict?
Why?

→ what might be the function of prediction?

- allows faster processing, increased efficiency

- reduced memory load

- to reduce ambiguity inherent in most linguistic utterances
Why?

⇒ The 'prediction is needed for learning' argument

- predictive dependencies (such that the or a is usually followed by a noun) allow language learners to acquire abstract structure (Saffran, 2010; Chang et al., 2006)

- the ability to extract statistical regularities is thought to be linked to an individual's prediction skills (cf. Conway et al., 2010; Misyak et al., 2010).

but:

- ability to extract forward statistical regularities does not necessarily tell us about the extent to which such results are driven by prediction

- “statistics of word sequencing are sometimes symptoms of meaning relations, but they do not constitute meaning relations” (Jackendoff, 2007)
Why?

→ The 'prediction is needed for learning' argument

- infants (Pelucchi et al., 2009) and adults (Perruchet and Desaulty, 2008) also track backward transitional probabilities (often more informative than forward transitional probabilities, St Clair et al. (2009)

- backward TPs are more informative than forward statistics to learn which of the articles in German (i.e. der, die, or das) precedes a noun since the articles itself are only poor predictors of a specific noun

→ language learning can take place in the absence of prediction since backward TPs can’t be used for prediction
Why?

The 'prediction is needed to coordinate dialogue' argument

- dialogue facilitates the alignment of interlocutors’ mental states (Pickering & Garrod, 2004) and explains the tight coupling of interlocutors

- prediction and imitation, Pickering and Garrod (2007) argue, explains why conversation typically is so effortless even though it involves constant task-switching and requires the planning of when to speak and what to say

- people often complete other's utterances (e.g., H. Clark & Wilkes-Gibbs, 1986)

- De Ruiter, Mitterer, & Enfield (2006): listener are highly accurate in predicting the end of other speakers' turns

- McFarland (2001): speakers coordinate their breathing

- Pardo (2006): conversation partners' pronunciations converge
Why?

→ The 'that's simply the way the mind works' argument

- Clark (2013): the action-oriented predictive processing framework has put forward a set of deep unifying principles for our understanding of the human mind in terms of neural function and organization

Friston (e.g., 2010):
- predictive coding involves the minimizing of prediction error though recurrent or reciprocal interactions among levels of a cortical hierarchy
What?

→ cues of predictions

ERP studies:
- sentence context (Federmeier & Kutas, 1999)
- discourse context (van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005)

Visual world eye-tracking:
- verb-selectional restrictions (Altmann & Kamide 1999)
- combinatorial information of subject and verb (Kamide, Altmann, & Haywood, 2003)
- discourse context (Kaiser & Trueswell, 2004; Altmann & Kamide, 2009)
- case-marking (Kamide, Scheepers, Altmann, 2003)
Altmann & Kamide (1999): verb-selectional information

→ Selectional information conveyed by the verb can be used to predict upcoming theme
What?
⇒ contents of predictions

ERP studies:
• semantic/conceptual features (Federmeier & Kutas, 1999; Federmeier, McLennan, De Ochoa, & Kutas, 2002)
• morphosyntactic features (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2003, 2004)
• phonological form (DeLong, Urbach & Kutas, 2005)
• orthographic form (Laszlo & Federmeier, 2012)

Visual world eye-tracking:
• semantic/conceptual features (Altmann & Kamide 1999)
• syntactic structures (Arai & Keller, 2012)

How far does pre-activation of information go?
Rommers, Meyer, Praamstra, & Huettig (2013, Neuropsychologia)

‘In 1969 Neil Armstrong was the first man to set foot on the moon’.

Target

Shape competitor

Control
Rommers, Meyer, Praamstra, & Huettig (2013, Neuropsychologia)

Fig. 2. Results of Experiment 1. Time-course graph showing fixation proportions to Targets, Shape competitors, and Control objects (solid lines) along with fixation proportions averaged across the three corresponding unrelated distractors (dashed lines). Display onset was at $-500$ ms, time zero indicates critical word onset.

→ visual form information can be predicted
Rommers, Meyer, Praamstra, & Huettig (2013, Neuropsychologia)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>In 1969 zette Neil Armstrong als eerste mens voet op de maan. In 1969 Neil Armstrong was the first man to set foot on the moon.</td>
</tr>
<tr>
<td>Shape</td>
<td>In 1969 zette Neil Armstrong als eerste mens voet op de tomaat. In 1969 Neil Armstrong was the first man to set foot on the tomato.</td>
</tr>
<tr>
<td>Unrelated</td>
<td>In 1969 zette Neil Armstrong als eerste mens voet op de rijst. In 1969 Neil Armstrong was the first man to set foot on the rice.</td>
</tr>
</tbody>
</table>

Fig. 4. Scalp topographies of the mean difference between each of the conditions (indicated by the two relevant waves taken from the Cz electrode): Unrelated—Correct, Shape—Correct, and Unrelated—Shape. Two time windows are shown. White dots indicate electrodes included in a significant cluster.

→visual form information can be predicted
How? Mechanisms

Dumb and smart routes to prediction: two-systems accounts

Kahneman’s (2011)
- system 1 operates “automatically and quickly, with little or no effort and no sense of voluntary control”

- system 2 is assumed to “allocate attention to the effortful mental activities that demand it, including complex computations … often associated with the subjective experience of agency, choice, and concentration”

- similar to Duncan’s (2010) multiple-demand system involving the frontal lobes

Huettig (2015):
- system 1 is the “dumb” route to prediction: simple associative mechanisms (e.g. based on Hebbian learning) lead to pre-activation of linguistic input (cf. Bar, 2007, 2009).

- system 2 is the “smart” route to prediction and linked to more effortful active reasoning
How? Mechanisms

Dumb and smart routes to prediction: two-systems accounts

Kuperberg (2007)

- has linked the N400 and P600 event-related potentials to two competing neural processing streams

- 1\textsuperscript{st} stream computes “semantic features, associative relationships and other types of semantic relationships between content words within a sentence, and compares these relationships with those that are pre-stored within lexical semantic memory” (N400 is sensitive to these computations)

- 2\textsuperscript{nd} processing stream is “combinatorial” and involves “the combination of words through algorithmic mechanisms to build up higher-order meaning” (P600 ERP component reflects a continued analysis within this combinatorial system)
How? Mechanisms

Dumb and smart routes to prediction: two-systems accounts

system 1:
- semantic priming studies suggest that temporal and inferior prefrontal regions are sensitive to associative relationships (e.g. Copland et al., 2003; Kotz et al., 2002; Matsumoto et al., 2005; Rossell et al., 2003)

- same brain regions (in fMRI studies) appear to be activated for the semantic incongruities that evoke the N400 in ERP studies (Hagoort et al., 2004; Kiehl et al., 2002; Kuperberg et al., 2003)

system 2:
- thought to involve posterior inferior frontal cortices, motor and parietal cortices, and middle and superior prefrontal cortices (Friederici et al., 2003; Kuperberg, 2007; Kuperberg et al., 2003; Ni et al., 2000).

- these regions appear to be activated by both morphosyntactic violations and semantic-thematic violations (and evoke P600s) consistent with the idea that system 2 is a “combinatorial stream”, sensitive to multiple linguistic constraints and the building up of higher-order meaning
Testing the multiple systems account

Hintz, Meyer, & Huettig (in prep) Exp. 1:
- participants read two sentences, establishing an event scenario
- followed by a final sentence containing one of three target words: a highly expected word, a semantically unexpected word that was related to the described event, or a semantically unexpected and event-unrelated word (cf. Metusalem et al., 2012)

*The parents were very excited about their new baby girl. One of the first things they did was to get her baptized in their church. The baby liked baths, so she smiled when she was sprinkled with WATER/PRIEST/DENTIST on her forehead.*
Testing the multiple systems account

Hintz, Meyer, & Huettig (in prep), Exp. 1:

*Grand average ERPs at the midline parietal electrode (Pz)*
Testing the multiple systems account

Hintz, Meyer, & Huettig (in prep) Exp. 2:
- associatively related words appeared in sentences which did not build up a coherent discourse context

*That Peter and Claudia had the boy baptized, was the subject of the conversation. Time and again the image of the church in the history book appeared to get his attention. The baby liked baths, so she smiled when she was sprinkled with WATER/PRIEST/DENTIST on her forehead.*
Testing the multiple systems account

Hintz, Meyer, & Huettig (in prep), Exp. 2:

Grand average ERPs at the midline parietal electrode (Pz)
Testing the multiple systems account

Hintz, Meyer, & Huettig (in prep):

Scalp topographies of the mean difference between the event-unrelated (red) and event-related (orange) conditions (indicated by two waves taken from the Pz electrode) in Experiment 1 (left) and Experiment 2 (right).
How? Mechanisms

Production-based prediction accounts

- assumes that we use the language production system covertly to anticipate language input (e.g., Chang et al., 2006; Dell and Chang, 2014; Pickering and Garrod, 2007, 2013)

Pickering & Garrod (2013):
- language users use forward production models in a similar way that actors use forward action models (cf. Wolpert et al., 2003)

- speakers are assumed to construct efference copies of their predicted productions and compare these copies with the output of a production implementer

- listeners also use these forward production models and covertly imitate speakers to predict the speaker's upcoming utterances

- these “predictions-by-simulation” are assumed to be impoverished representations rather than fully implemented production representations.
Toddlers’ anticipatory eye movements & production abilities (Mani & Huettig, 2012)

- preferential looking paradigm

Onset of images (0ms)

Pre-verb window
2000 – 3000ms

Verb window
233ms after verb onset until noun onset

Noun window
233 to 2000ms from noun onset

Trial end (8000ms)

Der Junge isst/sieht den großen Kuchen
The boy eats/see the big cake


→ toddlers prediction ability correlates with production vocabulary size but not with comprehension vocabulary size

→ consistent with production-based prediction
How? Mechanisms
Production-based prediction accounts

Huettig (2015):

- comprehenders sometimes use their language production system to anticipate what another person is likely to say

- people use their fully-fledged production system for this rather than making use of a forward model
How? Mechanisms
Prediction via event simulation

Tversky & Kahneman (1973):
- people predict the likelihood of an upcoming event by how easy it is to simulate it

Moulton & Kosslyn (2009):
- all mental imagery serves simulation
- imagery allows us to generate specific predictions based upon past experience

Huettig (2015): we often use event simulation heuristic (i.e. imagery) to pre-activate linguistic representations

How could imagery link up with linguistic representations?

- McQueen and Huettig (2014): seeing visual objects primes the semantic and phonological representations of related spoken words

→ no reason to believe that mental imagery of objects or events cannot do the same
How? Mechanisms
PACS – a multiple mechanisms account of prediction in language processing

Huettig (2015):
1) upcoming input is pre-activated by simple associative mechanisms (e.g. Hebbian learning)

2) pre-activation of linguistic input via combinatorial mechanisms which are sensitive to multiple linguistic constraints and the building up of higher-order meaning

3) comprehenders sometimes use their fully-fledged language production system to anticipate what another person is likely to say

4) linguistic input can be activated via event simulation

→ mechanisms which are minimally required for a comprehensive account of predictive language processing
How? Mechanisms
PACS – a multiple mechanisms account of prediction in language processing

What is the nature of the relationship between these different mechanisms?
- mechanisms are not encapsulated but constantly interact

- associations feed into combinatorial mechanisms
  \( \rightarrow \) associative priming leads to (faster) retrieval of associated representations/words

- production system uses these combinatorial mechanisms but crucially there is more to language production than sequential combinatorial operations

- context provides the trigger for the use of the fully fledged production system for prediction (e.g. when people complete each other's utterances)

- event simulations may trigger associations between particular types of events and particular representations/words/syntactic structures and vice versa

- event simulation may be more likely to occur during natural conversation thereby providing possible associative links between event simulation and production-based prediction
How? Mechanisms
PACS – a multiple mechanisms account of prediction in language processing

Why would our minds work that way?

- evolutionary advantages for minds who can predict upcoming linguistic input by making use of multiple routes/mechanisms?
  → such minds may have been able to draw behavioural consequences more rapidly than minds employing the same old strategy in every context/situation

- no evolutionary benefit?
  → evolution serves no purpose

→ multiple mechanisms in predictive language processing may simply reflect evolutionary history
How?
Mediating factors of predictive language processing

1) literacy (Mishra, Singh, Pandey, & Huettig, 2012; Mani & Huettig, 2014; Huettig & Brouwer, 2015)

2) working memory & processing speed (Huettig & Janse, in press)

3) age (Federmeier et al., 2010; Huang et al., 2012; Wlotko and Federmeier, 2012; but Huettig & Janse, 2012)
When?
Do people always predict (cf. Clark, 2013; Friston, 2010)

- most studies on predictive language processing have used very high cloze probability sentences

- experimental evidence that language comprehenders do not always predict (Mishra et al., 2012; Mani & Huettig, 2014; Federmeier et al., 2012)
→ strong need for more diverse participant populations

→ cross-linguistic study is required (only a tiny number of the world's thousands of languages have been explored; languages differ dramatically at all levels of linguistic organization, Evans & Levinson, 2009)

→ psychological theories should be evaluated using individual differences (Kosslyn et al., 2002; Vogel & Awh, 2008)

→ use of diverse experimental paradigms (methods of choice such as EEG and visual world eye-tracking have important strengths but also their own particular limitations)
Do people always predict (cf. Clark, 2013; Friston, 2010)

- current evidence also consistent with the notion that prediction is an important aspect but not a fundamental principle of language processing
Review

Four central questions about prediction in language processing

Falk Huettig\textsuperscript{a, b, *}

\textsuperscript{a}Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
\textsuperscript{b}Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands