

Bachelor Thesis  
Artificial Intelligence



Radboud University Nijmegen

---

Capturing interactive alignment with the  
predictive processing framework

---

Author:  
Gerardo Maio

Supervisor:  
Dr Mark Blokpoel  
Department of Artificial Intelligence  
Faculty of Social Sciences  
Radboud University Nijmegen

05 July 2017

Abstract

The predictive processing theory is believed to be able to capture all cognitive processes and therefore should also be able to capture alignment phenomena in communication. The interactive or integrated alignment theory describes how alignment can work when interlocutors are in easy dialogue and is a very promising theory of human cognition. Thus, if interactive alignment is indeed a part of human cognition, the predictive processing framework should be able to capture interactive alignment phenomena. This thesis shows how interactive alignment can be described using predictive processing and what the benefits and perks of this approach are. The predictive processing framework can help to broaden the interactive alignment theory so that it can describe more than just very easy dialogues and it can contribute to the development of a full computational implementation of the interactive alignment theory, which is not yet available. Although predictive processing can describe basic alignment, there are further challenges, such as capturing communication with its multimodality and in all its different forms.

*Keywords: predictive processing, interactive, integrated, alignment, dialogue*

**Table of contents**

Abstract ..... 2

Capturing interactive alignment with the predictive processing framework ..... 5

Predictive processing framework ..... 8

    Explanation..... 8

        Prediction Error. .... 10

            Update hypotheses..... 11

            Update causal model. .... 11

            Gather evidence..... 11

            Active inference and lowering the level of detail. .... 12

        Predictive Processing in a nutshell..... 12

Communication ..... 13

    Interactive alignment..... 14

        Alignment mechanisms..... 15

            Priming..... 15

            Percolation..... 17

            Repair. .... 17

            Routines..... 18

            Monitor..... 18

        Predictions..... 19

    Summary ..... 19

    Remarks on the interactive alignment theory..... 20

Capturing interactive alignment with predictive processing ..... 22

    Levels of knowledge ..... 22

    Production and comprehension ..... 23

        Production. .... 23

        Comprehension..... 24

Alignment mechanisms .....	27
Priming in predictive processing.....	27
Repair in predictive processing. ....	27
Update hypothesis. ....	28
Update causal model. ....	28
Lowering the level of detail. ....	28
Gather evidence.....	29
Active inference. ....	29
Routines.....	30
Monitor.....	32
Discussion .....	34
Acknowledgements .....	38
References .....	39

### Capturing interactive alignment with the predictive processing framework

Communication is a core skill of human beings, and it is ubiquitous in our society, which makes it a significant field of research. Some theories assume that one considers information about the addressed when creating and expressing a new message (Bell, 1984; Sacks, Schegloff, & Jefferson, 1974). To communicate with each other humans use speech and gestures, and in order to connect and understand one another, interlocutors are continuously making predictions about the addressed person (H. H. Clark, 1996). One theory about this phenomenon is called audience design and, according to Allan Bell (1984), its most important message is that speakers are interactively designing their style of speech for their audience and in response to it (Eckert & Rickford, 2001, pp. 139-169). With this statement Eckert and Rickfort (2001) claim that humans make use of the theory of mind, i.e. they know that their knowledge is different to the knowledge of others, which sets humans apart from many other animals (Frith & Frith, 2005; Tomasello, 2009). Some studies have shown that some alignment must be happening in communication and they demonstrate the importance of understanding this phenomenon. For example Galati and Brennan (2014) found that interlocutors take into account which events are known to the counterpart and coordinate that information, which is then translated into speech and gestures (Galati & Brennan, 2014). Moreover, Friston and Frith (2015) wrote a paper about active inference, wherein they propose that interlocutors predict each other's actions, which helps them to understand one another and to align their messages so that the addressed will be able to understand the utterance (K. J. Friston & Frith, 2015). By understanding alignment in communication, researchers can bring that knowledge to a higher level and try to generalise it to other parts of human cognition. If the alignment mechanisms are true in communication, they could be holding for other domains too.

The most elaborate theory on alignment in communication is developed by Pickering and Garrod (2004), who concluded that interactive alignment can be seen as a crucial part of successful communication between interlocutors (Garrod & Pickering, 2013; Pickering & Garrod, 2004). While the interactive alignment theory is an innovative and promising theory, there are “no full implementations” available at the moment (Wachsmuth, De Ruiter, Jaecks, & Kopp, 2013). Hopefully, predictive processing can help to realise that and perhaps tackle some critics that

erected in the past. In this thesis, it will be tried to find and possibly correct flaws in the interactive alignment theory by modelling it with predictive processing. Of course, it is also possible to find challenges for the predictive processing framework which will need to be addressed.

The just mentioned predictive processing framework is assumed to be able to capture all human cognition. Clark states, that the “hierarchical prediction machine” is the best “unifying model of perception and action” (A. Clark, 2013; H. Clark, 1996). Assuming that the interactive alignment theory of Pickering and Garrod (2004) do capture the cognitive phenomenon of alignment correctly, and considering the unifying character of predictive processing, the predictive processing framework should be able to capture these alignment phenomena. The predictive processing framework assumes that we make predictions about everything we perceive and that we try to minimise the prediction error. Thus human cognition is minimising the error between the observed input and the prediction made by “the generative model and inferred causes” (K. Friston, 2005). This error minimization is done by using different computational processes which will be explained later on (Kwisthout, Bekkering, & van Rooij, 2016; Thornton, 2016b). The predictive processing framework gets more and more attention as it explains our brains as “multilevel prediction engines”, which seems to be in accordance with theories of cognitive psychology and psychophysics (A. Clark, 2015a). Thornton (2016a) showed that predictive processing could be an explanation of how human brains compute and thus how cognitive processes in humans work (Thornton, 2016a).

Because of the dissertations above, it is evident, that the predictive processing framework must be further investigated, as it seems to be a very promising view of cognition. This thesis tries to capture the interactive alignment theory using the predictive processing framework. As this framework should unify all cognition, it should be able to model communication as a whole. Therefore it must also be able to capture alignment occurrences as described in the interactive alignment theory, which is only a fraction of communication. Through the entire thesis it is assumed that the interactive alignment theory of Pickering and Garrod (2014, 2013) captures alignment phenomena in communication. If the framework fails in capturing these alignment processes, one must either change the perspective on predictive processing as a unifying theory of cognition or human communication

as interactive alignment processes described by Pickering and Garrod. The latter form of dealing with this problem would also lead to the question if alignment in human cognition can be assumed. The underlying question is: Can the predictive processing framework capture interactive alignment in communication? To answer this question, some explanation steps are needed. First of all, the predictive processing framework needs to be made very clear. Due to the rather confusing use of the terms predictive processing, predictive coding and radical predictive processing, this thesis gives an overview of the definition that is used in this project. The next step is to characterise and define the interactive alignment theory. Finally, these theories will be brought together to see if the predictive processing framework is able to capture interactive alignment and to track down problems in doing so. Such problems can be used to improve one or both theories. It could, for example, be possible that predictive processing can help to come closer to an implementation of the interactive alignment theory.

### **Predictive processing framework**

As mentioned before the predictive processing framework is believed to be a unifying theory of cognition and should thus be able to capture all cognitive processes of human beings (A. Clark, 2013; H. H. Clark, 1996). The theory assumes that people make predictions of the world and that these predictions are then verified or falsified (A. Clark, 2015a). If a prediction is verified with the real world input, then there will be no error and nothing else happens (A. Clark, 2013). If a prediction is falsified, it needs to be updated, as it does not fit the real world input (A. Clark, 2013). This updating happens in small steps, as there already was evidence for the prediction in the first place. Hence it would not be smart to entirely replace the prediction (A. Clark, 2013; Kwisthout et al., 2016). The model one has over the world will therefore slowly converge to the real world situations one experiences. However, it is important to understand, that the predictions one has about the world are also influencing the real world indirectly through one's actions (Kwisthout et al., 2016). Clark (2015b) gives a compelling example of how predictions and the real world are affecting one another. He explains, that when a person thinks someone is angry at her, then she will perceive his face as angry the next time she sees him. Therefore she will align her behaviour, and he might think that something is wrong with her and act differently. Thus her predictions of the world affect her actions and as a result change the real world. When she finds out that he is not angry at her, her predictions are being updated. Hence predictions and the real world do affect each other through actions (A. Clark, 2015b).

The following chapters will provide some explanation of predictive processing.

#### **Explanation**

What is termed predictive processing in this thesis has many names in other articles. The framework is sometimes denoted as predictive processing (A. Clark, 2015a; Kanai, Komura, Shipp, & Friston, 2015; Kwisthout et al., 2016), predictive coding (Brown & Friston, 2012; Huang & Rao, 2011; Lupyan & Clark, 2015), predictive mind (Hohwy, 2013) or Bayesian brain (Doya, 2007; Knill & Pouget, 2004; Seth, 2014). All of them use the same fundamental framework, but add some more theory and ideas to it. In this thesis, the framework will be denoted as predictive processing, and this chapter will give a summary of the used framework.

The predictive processing framework works with top-down and bottom-up processes (Lupyan & Clark, 2015). It is assumed that predictions are quite abstract

and that they are made on a higher level of cognition (Lupyan & Clark, 2015). In fact, the higher the level is the more abstract the predictions get. These predictions thus come from a high-level top-down to the predictions that are very close to the actual input, which means that the low-level predictions are highly spatially and temporally precise (A. Clark, 2015b; Lupyan & Clark, 2015). On the other hand, there are bottom-up signals which are based on the information from the real world and one's predictions. Low-level predictions are being compared to the actual input, and if there is an error, this error will be passed on to higher level predictions, if the precision weighting allows that (Lupyan & Clark, 2015).

The precision of a prediction error is a measure which decides what is done with a certain prediction error by indicating the nature of the error (Kwisthout et al., 2016). Predictions do also have precision estimates, which capture the uncertainty of the prediction itself (Kwisthout et al., 2016). If one makes a whole new experience and thus lacks knowledge, then the prediction error will affect earlier predictions, and therefore the knowledge is broadened. If on the other hand, there is a natural chance that one's predictions in that situation are not always right, then the prediction error is just being ignored. The latter is the case when someone tosses a coin, as there is inherent uncertainty over the outcome of this toss: it can be heads or tails with a 50% chance on both, so there is no need to change the hypotheses (Kwisthout et al., 2016). These precisions are also just expectations of an agent and can, therefore, be updated too, if there is any reason to do so (Kwisthout et al., 2016). A prediction error which is expected to have a high precision will lead to an internal or external inference, whereas prediction errors with a low expected precision will not result in such alterations (Seth, 2014).

Figure 1 gives an overview of the predictive processing framework with bottom-up and top-down processes. Note that only the unpredicted input, i.e. prediction errors, is forwarded upwards, whereas correctly predicted information is not. Thus only prediction errors will be processed in a way that changes the hypotheses or model of an agent. The figure is adapted from G. Lupyan and A. Clark (2015). For the purpose of this thesis, the term 'precision gate' has been added as a yellow flash. It is used to show, that not all prediction errors will lead to an update of the hypotheses or model of an agent. When the assumed precision of a prediction error is too low, i.e. it does not hit a certain threshold, the prediction error will not be forwarded upwards the levels. The higher the precision, the more likely

the error is to change the cognitive model of an agent (Kwisthout et al., 2016; Lupyan & Clark, 2015; Seth, 2014).

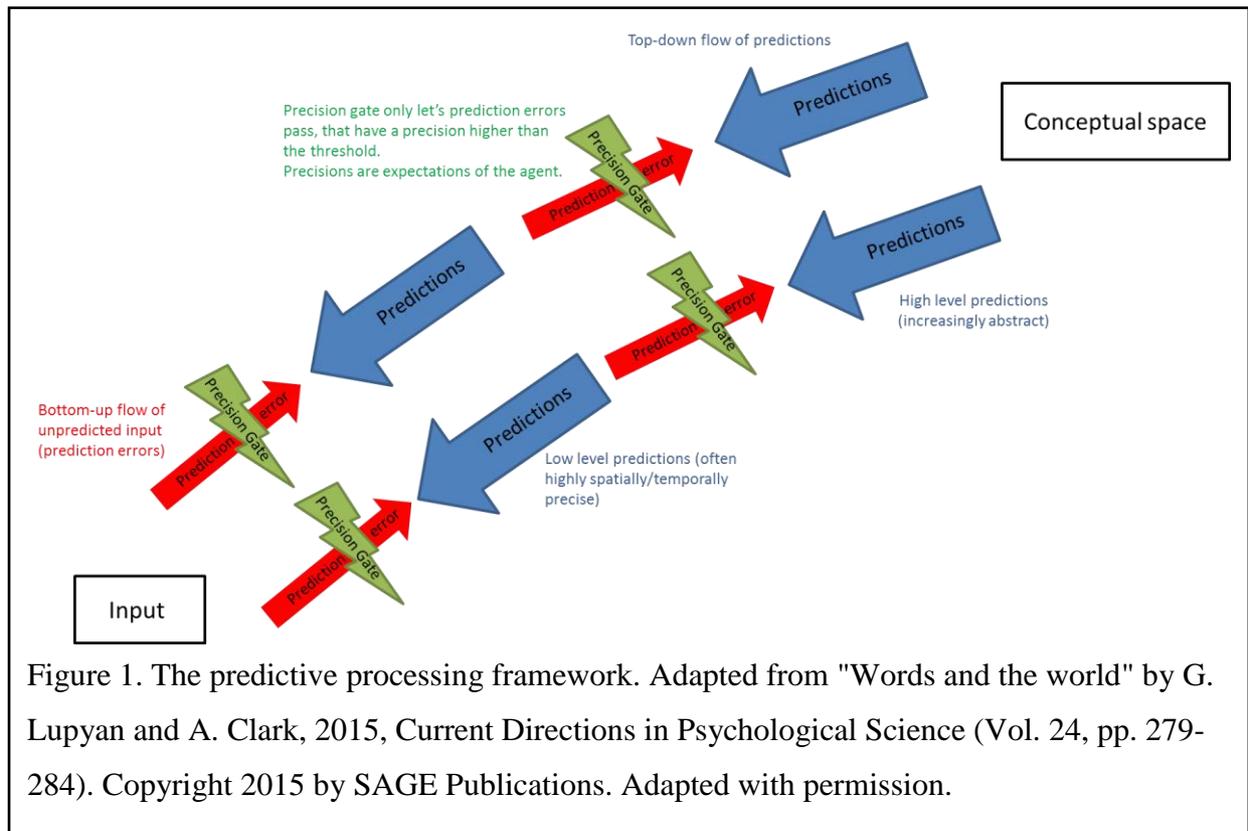


Figure 1. The predictive processing framework. Adapted from "Words and the world" by G. Lupyan and A. Clark, 2015, Current Directions in Psychological Science (Vol. 24, pp. 279-284). Copyright 2015 by SAGE Publications. Adapted with permission.

Most scientific articles and theories on predictive processing assume that it can be explained using Bayesian networks (A. Clark, 2015b; Kwisthout et al., 2016; Payzan-LeNestour & Bossaerts, 2011). Harkness and Keshva (2017) have shown that predictive processing and Bayesian networks can be very beneficial for each other and that they complement one another. They argue that predictive processing is a theory about how Bayes’ optimal behaviour happens. In this thesis, Bayesian networks will be used to describe predictive processing and finally to capture the interactive alignment theory.

**Prediction Error.** Prediction errors are here defined as the Kullback–Leibler divergence between the observed and the predicted distribution<sup>1</sup> (K. Friston, 2010; Thornton, 2016a). If this value is high, one needs to minimise it (K. Friston, 2010; Thornton, 2016a). Prediction errors can be understood as a value for surprise. If an agent is surprised by some event, then clearly the prior predictions were wrong

<sup>1</sup> The Kullback–Leibler divergence is denoted as  $D_{KL}(O||P) = \Pr_O(p) * \log\left(\frac{\Pr_O(p)}{\Pr_P(p)}\right)$  where ‘O’ and ‘P’ are the observed and predicted probability distributions respectively.

(Kwisthout et al., 2016). An event that matches the prediction will not lead to surprisal and thus will not be seen as a prediction error (Kwisthout et al., 2016). The average of this variable can also be denoted as free energy (Seth, 2014). Kwisthout et al. (2016) described several ways to deal with prediction errors, which will be summarised in the following sections.

***Update hypotheses.*** When prediction errors occur, these are processed in a bottom-up way through the predictions, and it might become necessary to update the hypotheses. This updating is not a change of the complete model but only of “the distribution over the hypothesis variables” and is used when the statistics are known and the real world input is possible but rather unexpected (Kwisthout et al., 2016). The posterior probability is calculated so that the prior probability can be updated for future predictions (Kwisthout et al., 2016).

***Update causal model.*** Sometimes it is necessary to update the causal model or its probabilities. In particular, this is needed when the uncertainty of a situation is unexpected; thus there is a lack of knowledge, a misrepresentation or a change in the real world (Kwisthout et al., 2016). Kwisthout et al. (2016) points out, that there are several ways to update the model. The update can be done by complementing the model with new variables, changing the values that can be attributed to the variables and by revising the conditional probability distributions (Kwisthout et al., 2016). Unfortunately, it is not yet described when which of the just mentioned methods are used and how an agent chooses which one to use.

***Gather evidence.*** Another approach to handle prediction errors is to gather more evidence so that one can decide which hypothesis is more likely, which, in turn, will lead to lower prediction errors (Kwisthout et al., 2016). One might argue that this method has its problems because it will result in biased predictions. However, as the predictive processing framework is meant to capture cognitive processes and biasing is found in human cognition, the framework is just in accordance with human cognition. Think of the example where one person is angry with another. Gathering evidence will lead to real-world errors, but internal prediction errors are reduced as the real world is being determined by one's predictions (A. Clark, 2015b). This phenomenon is widely investigated in the field of psychology and is denoted as confirmation bias (Klayman, 1995). People often tend to search only for evidence that supports the preferred hypothesis (Klayman & Ha, 1987). This is done by ignoring falsifying aspects, gathering evidence only in

accordance with the hypothesis and overweighting or underweighting some experiences (Klayman & Ha, 1987). Hence, the mechanism of gathering evidence can be used to handle prediction errors.

*Active inference and lowering the level of detail.* Furthermore one can handle prediction errors by affecting the real world so that it matches the own hypothesis or by using more abstract predictions so that prediction errors can be prevented (Kwisthout et al., 2016). In active inference, one is engaging in the world to change it actively. If one predicts a thing to be in a certain position, it is possible to change the prediction, but one can also just modify the position of the object in the real world. An example of lowering the level of detail is when one tosses a coin. One can try to predict whether it will show head or tails, but one could also hypothesise that the coin eventually will land somewhere no matter if it shows heads or tails. The second hypothesis is more abstract than the first one. The level of detail hence can be changed to minimise prediction errors.

### **Predictive processing in a nutshell**

Let's summarise the story very shortly. In predictive processing, the brain is seen as a hypothesis-testing-machine with several hierarchical levels of processing. The brain makes top-down predictions based on some hypotheses, which are then compared to the real-world observation and prediction errors, i.e. differences between the observation and the prediction, are being processed in a bottom-up fashion. The central aim of predictive processing is to minimise prediction errors using precision-weighted predictions. By minimising the average of prediction errors in the long-run, humans try to understand the world and make better predictions in the future. Agents can either change the observation (active inference) or change the internal models. Bayesian inference helps to decide how an agent should update his model when an unpredicted input is presented.

This concludes the chapter on predictive processing. As this thesis tries to capture the interactive alignment theory of communication with the predictive processing framework, the next chapter will give an overview of the interactive alignment account before proceeding with the research question.

## Communication

According to the Oxford Dictionary communication is the “imparting or exchanging of information by speaking, writing, or using some other medium” (“Communication,” 2017). This definition is quite general, and of course, there are more definitions and theories about communication, as will be discovered in this chapter of the thesis (Stevens, 1950). However, this definition already demonstrates that communication is ubiquitous in our society. Humans are constantly communicating with one another. Because of its core role in human society, there are many theories about communication.

The most famous theory is the Shannon-Weaver model, also denoted as the code model. This model suggests that there is a sender, who makes a message and uses a channel to send this message to someone (Shannon, 2001). Before using the channel, the sender must encode the message (Shannon, 2001). After sending the message through the channel, where some noise, like environmental influences, will be added to the message, the message needs to be decoded again, so that the receiver can understand it (Shannon, 2001). The receiver then provides the sender with feedback and therefore gets to be the sender himself (Shannon, 2001). As famous as this model may be, there has been a lot of critics on this theory. Chandler (1994) summarises the critics on it in one article. He criticises that interlocutors cannot be separated into sender and receiver, as they communicate simultaneously with speech and body. Thus the linear model is not realistic. Furthermore, the role of the content of a message and the interpretation of that message is not taken into account (Chandler, 1994). Chandler states that the receiver must interpret the message, which makes him not a passive receiver, but an active interpreter of the message. The lack of context and relationships in this theory is yet another critic of Chandler and others (Chandler, 1994; Garcia, 1996).

It becomes apparent that this theory cannot satisfactorily be used to explain human communication as it is. New theories arose, but as Wachsmuth et al. state there is no theory of communication that can “cover the multitude of observed communicational phenomena” and is “concrete enough to provide guidelines . . . for implementing communicative systems in artificial agents” (Wachsmuth et al., 2013).

Pickering and Garrod (2004) made a new proposal on how to understand and implement communication, which seems to be very promising. They developed the

interactive alignment account, which sets the basis of alignment theory in communication (Pickering & Garrod, 2004). The authors have been working on their theory ever since and added some concepts to it later on (Pickering & Garrod, 2007, 2013). The more developed theory is denoted as integrated alignment theory (Pickering & Garrod, 2013). A combination will be discussed in the following chapter.

### **Interactive alignment**

Pickering and Garrod (2004) wanted to show how successful dialogue between interlocutors works and used alignment to do so. They defined alignment as interlocutors sharing the same representation on one or more particular levels, which are defined as linguistic levels, the level of the situation model and the level of general knowledge. Situation models are mental representations of the events that are being discussed, and according to the interactive alignment theory, these situation models will align during a conversation (Pickering & Garrod, 2004). The mental representations do include causal, personal, spatial and temporal information of the situation in mind (Zwaan & Radvansky, 1998). Johnson-Laird (1980) stated that a situation model is absolutely necessary for communication and understanding, because of its nature of being a representation of the meaning of words. Pickering and Garrod (2004) argue that if the situation model of the interlocutors is not the same, they may end up talking about entirely different events and will not understand each other as intended. Therefore, according to them, alignment of the situational model is needed, so that both participants will know what they are talking about eventually. Thus it can be assumed, that a general understanding of the discussed situation is needed to communicate with one another successfully. This understanding can be achieved by alignment of the situation model.

Pickering and Garrod (Pickering & Garrod, 2004, 2006) explained that alignment on other cognitive levels is needed to end up aligning the situation model finally. They defined three different levels of knowledge: general knowledge, situation model and linguistic knowledge. Several studies have shown that interlocutors do align their linguistic expressions to one another. For example Branigan et al. (2000) found that interlocutors tend to use the same syntactic structure. This phenomenon is called syntactic or structural priming and is found in comprehension and production of speech (Bock, 1986; Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995). They are coordinating their linguistic skills

and aligning their syntax to one another (Branigan, Pickering, & Cleland, 2000). Such alignment was also found on other linguistic levels such as expressions and words (Garrod & Anderson, 1987), accent and speech rate (Giles, Coupland, & Coupland, 1991) and articulatory level (Bard et al., 2000).

Pickering and Garrod (2006) defined general knowledge as a contrast to linguistic knowledge and the situation model. To have a situation model, one also needs general knowledge linked to that situation model (Pickering & Garrod, 2006). They describe that general knowledge is not a part of the situation model as a lot of that knowledge is irrelevant in that situation, but it is often involved when thinking about a situation. As the authors put it, some “knowledge is relevant to the situation model but is clearly not part of it”. Unfortunately they did not describe how an agent makes a distinction between relevant and irrelevant knowledge.

Now that alignment, general knowledge, situation model and linguistic knowledge are defined, it is possible to go into more detail about alignment and how the interactive alignment theory proposes the mechanisms to work.

**Alignment mechanisms.** In order to understand the mechanisms of the interactive alignment theory, there will be some references to Figure 2. This is a scheme proposed by Pickering and Garrod (2004) and is meant to capture alignment in dialogue. The authors define certain mechanisms that lead to alignment: priming across interlocutors, percolation between levels, reparation of misalignment, the use of routines and routinization, and monitoring.

**Priming.** According to Pickering and Garrod (2004), the most important mechanism involved in alignment is priming, and it is unconscious and automatic. In Figure 2 one can see arrows between the levels of both speakers. With these arrows, they want to represent priming between interlocutors. Suppose speaker A produces a sentence and when doing so, he activates a situation model and corresponding linguistic knowledge. To understand the message of speaker A, speaker B has to activate the same situation model and linguistic knowledge. As both interlocutors can comprehend and produce language, the arrows in Figure 2 go both ways. These arrows are called “alignment channels” and represent alignment between interlocutors on situational and linguistic levels (Pickering & Garrod, 2004). When two interlocutors have aligned, the certain situation models of both interlocutors “contain the same entities and relations”, and their linguistic

knowledge is aligned as “they have similar patterns of activation of linguistic knowledge” (Pickering & Garrod, 2004, 2006).

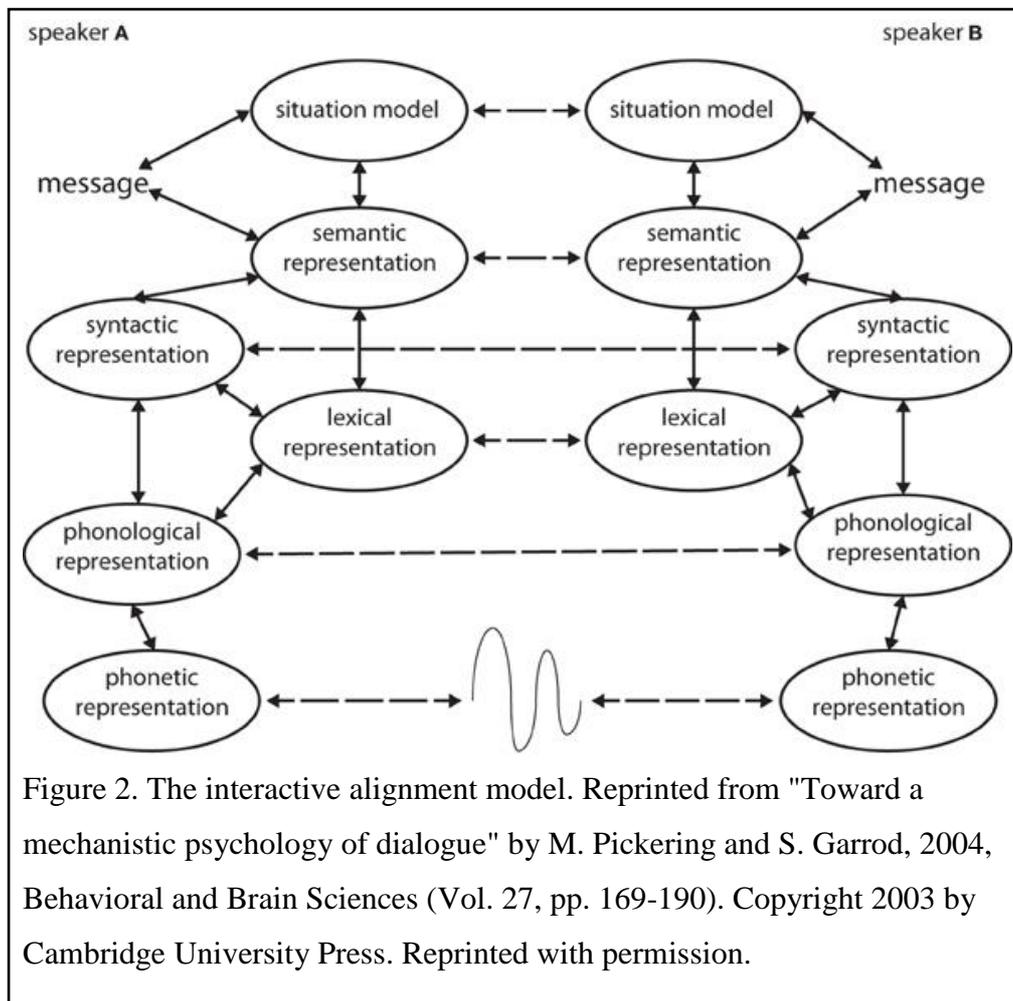


Figure 2. The interactive alignment model. Reprinted from "Toward a mechanistic psychology of dialogue" by M. Pickering and S. Garrod, 2004, Behavioral and Brain Sciences (Vol. 27, pp. 169-190). Copyright 2003 by Cambridge University Press. Reprinted with permission.

Because of alignment on the situation model, there can also be alignment of general knowledge (Pickering & Garrod, 2004). Although general knowledge is not used in Figure 2, Pickering and Garrod (2006) propose that this level is higher than the situation model. As alignment on the situation model is seen as the most important kind of alignment in communication, the authors do not implement it as a level in Figure 2. To activate a situation model one also uses some general knowledge that is linked to the situation model (Pickering & Garrod, 2006). Thus, when speaker A activates a situation model with some information from the general knowledge, the same activation is triggered in speaker B. However, Pickering and Garrod (2006) point out that alignment of general knowledge is slow and partial.

As already explained, the arrows between interlocutors go both ways. This shows that “parity between comprehension and production” is implied in interactive

alignment (Pickering & Garrod, 2004). A representation that is built to understand an utterance can therefore also be used for production and vice versa (Pickering & Garrod, 2004). Representations can be reused, but it is important to add that this parity does not imply the equivalence of comprehension and production processes (Pickering & Garrod, 2004).

**Percolation.** As depicted in Figure 2, there are also connections between the different levels. Pickering and Garrod (2004) have argued that alignment on one level leads to alignment on other levels as well. They explain that interlocutors aligning on a particular level, for example on the lexical level will also align on other levels – in this case, alignment is enhanced on a syntactic level. Thus, interlocutors “tend to align expressions at many different levels at the same time” (Pickering & Garrod, 2004).

**Repair.** To understand the repair mechanism, one must first know how implicit common ground and full common ground are defined in the interactive alignment model. Pickering and Garrod (2004) defined implicit common ground as shared information of both interlocutors. They point out that the shared information is automatically built through alignment over time and thus is not a result of explicit modelling. The interlocutors comprehend each other’s utterances and make the changes on their situation model because they correct their internal representation to match with the utterance (Pickering & Garrod, 2004).

When the implicit common ground is erroneous interlocutors need to repair it interactively (Pickering & Garrod, 2004). To do so, one first needs to examine if it is possible to “interpret the input in relation to one’s own representation” and if that fails, one needs to reformulate “the utterance in a way that leads to the establishment of implicit common ground” (Pickering & Garrod, 2004). This kind of reparation can only happen interactively as the interlocutors need to talk to each other to fix the misalignment (Pickering & Garrod, 2004). However, there is no use of the listener’s situation model when one is reformulating a sentence. A speaker is only investigating the conversation “in relation to his own knowledge of the situation” and not considering the counterparts knowledge (Pickering & Garrod, 2004).

It is also possible to repair misalignment using full common ground. This means that a speaker has to make inferences about the listener. Examples of this process are explicit negotiations and discussions of the situation model. As this

mechanism is very costly, it is not part of easy conversations and is not being discussed in detail (Pickering & Garrod, 2004). However, hopefully, it is possible to also model repair using full common ground with the predictive processing framework so that the interactive alignment theory can broaden its scope to more than just easy dialogue.

**Routines.** Pickering and Garrod (2004) define routines as fixed expressions that have a higher frequency than other expressions. There are some very stable routines, such as ‘How are you?’, but according to the interactive alignment theory, there are also routines that are set up during dialogue and only for the sake of this particular dialogue. This process is called routinization (Pickering & Garrod, 2004). Suppose a dialogue between two interlocutors where one uses a particular expression which is not quite common in that situation. For the sake of the dialogue, the other interlocutor will probably use the same expression to keep the conversation flowing smoothly. After this dialogue, both interlocutors may give up the expression and thus not use it again to describe that situation (Pickering & Garrod, 2004). Of course, it is possible that such a routine will be used again after this particular conversation (Pickering & Garrod, 2006). After the routine is developed, users do not need to build up the expression with its linguistic components and its meaning again (Pickering & Garrod, 2004). It is short- or long-term stored in one’s lexicon, which makes it easier to produce and comprehend these expressions (Pickering & Garrod, 2004). The routine is likely to be long-term stored when the expression is used very frequently, i.e. repeated quite often (Pickering & Garrod, 2004).

**Monitor.** According to the interactive alignment model interlocutors are affected by the representations of the counterpart in the conversation (Pickering & Garrod, 2004). This effect is explained with priming and is represented by the horizontal arrows between the levels of both interlocutors in Figure 2. Due to this effect, Pickering and Garrod (2004) assume that “a speaker can also be affected by his own representations at these levels”. They stress that in self-monitoring one does not need to hear his own utterance, but that monitoring is internal and automatic on the different levels. A speaker is constantly monitoring his own contributions “with respect to the implicit common ground”, while a listener is monitoring her “partner’s contributions with respect to the same implicit common ground” (Pickering & Garrod, 2004).

**Predictions.** Later on, Pickering and Garrod (2007, 2013) added predictions to their theory and gave a clear distinction between language production and language comprehension. They assume that production is a top-down process from message to phonetics and that comprehension is a bottom-up process from phonetics to message. Pickering and Garrod (2007) propose that comprehending interlocutors predict what the other one is going to say with their own production system. This prediction activates representations on all levels and therefore will act as a prime in the production system of the listener, who is going to use related linguistic structures (Pickering & Garrod, 2007). The authors describe that producing speakers predict their utterances and compare these predictions to the output of their production system. Note that Pickering and Garrod (2007, 2013) propose that a listener does not immediately predict the situation model, but only the linguistic levels.

### **Summary**

As the alignment theory is quite complex, this section will give a concise overview of the theory. Pickering and Garrod (2004, 2006) propose that interlocutors construct similar situation models when they are in a successful dialogue. In order to align on the level of the situation model, one first needs to align on linguistic elements such as phonetics, phonology and so forth (Pickering & Garrod, 2004). Alignment mostly happens as a result of priming mechanisms, and because of it, interlocutors automatically construct an implicit common ground representing the shared information (Pickering & Garrod, 2004). When misalignment does occur, it must be repaired using implicit common ground or full common ground (Pickering & Garrod, 2004).

In their theory, there are also routines, which are fixed expressions with a high frequency and users monitor their own contributions without hearing the utterance itself, to account for errors. Monitoring is an internal process and can lead to alignment on some levels (Pickering & Garrod, 2004, 2006).

According to Pickering and Garrod (2013), interlocutors use predictions to communicate. They assume that in language comprehension, users imitate the so far uttered sounds, derive their message, generate efference copies and then compare these copies with the utterance of the speaker, which is, in fact, the input. In language production, users construct efference copies of the predicted utterance and compare these with their own internal output of the levels (Pickering & Garrod,

2013). They also point out, that language comprehension and language production are interwoven processes. In other words, during comprehension one uses the own possible next utterances and predicts what the other one is going to say and during production, one uses comprehension possibilities to predict if an utterance is understandable. Users want to “get ahead of the game” using these mechanisms (Pickering & Garrod, 2013).

### **Remarks on the interactive alignment theory**

To complete this chapter a few critical remarks are needed. During the development of this thesis, some questions concerning the theory of Pickering and Garrod and the differentiation of some theories arose, which will be shortly answered in this part of the chapter.

First, it is necessary to stress that Pickering and Garrod (2004) developed their alignment theory to only capture easy dialogue (Pickering & Garrod, 2004). They assume dialogue to be a joint activity of two interlocutors, who do cooperate to achieve the aim of that dialogue, which is the understanding of it in both interlocutors. When a parent tries to explain some difficult things to a child or when a teacher attempts to explain some challenging problems, this cannot be seen as an easy conversation in the sense of Pickering and Garrod.

Another relevant point is, that, in this thesis, the developments of Pickering and Garrod are summarised to one theory. The authors are constantly working on their model and added predictions in a later stadium (Garrod & Pickering, 2013; Pickering & Garrod, 2004, 2006, 2007, 2013).

Furthermore, it is important to notice that alignment is not a version of auditory design, recipient design or theory of mind. These theories assume that an interlocutor uses his knowledge about the other interlocutor to make utterances that the other one should understand (Bell, 1984; Leslie, 1987; Sacks et al., 1974). Thus one does not only use her own situation model but also considers the knowledge of the other interlocutor when communicating with someone (Bell, 1984; Leslie, 1987; Sacks et al., 1974). However, Pickering and Garrod (2004, 2006) do explain that this kind of alignment is way too costly to be involved in easy conversations. According to them, this technique of using the knowledge about the other one’s knowledge to produce a sentence is only used when one needs to repair misalignments. What looks like audience or recipient design can be seen as a result of proper alignment. As described in this chapter interlocutors build up an implicit

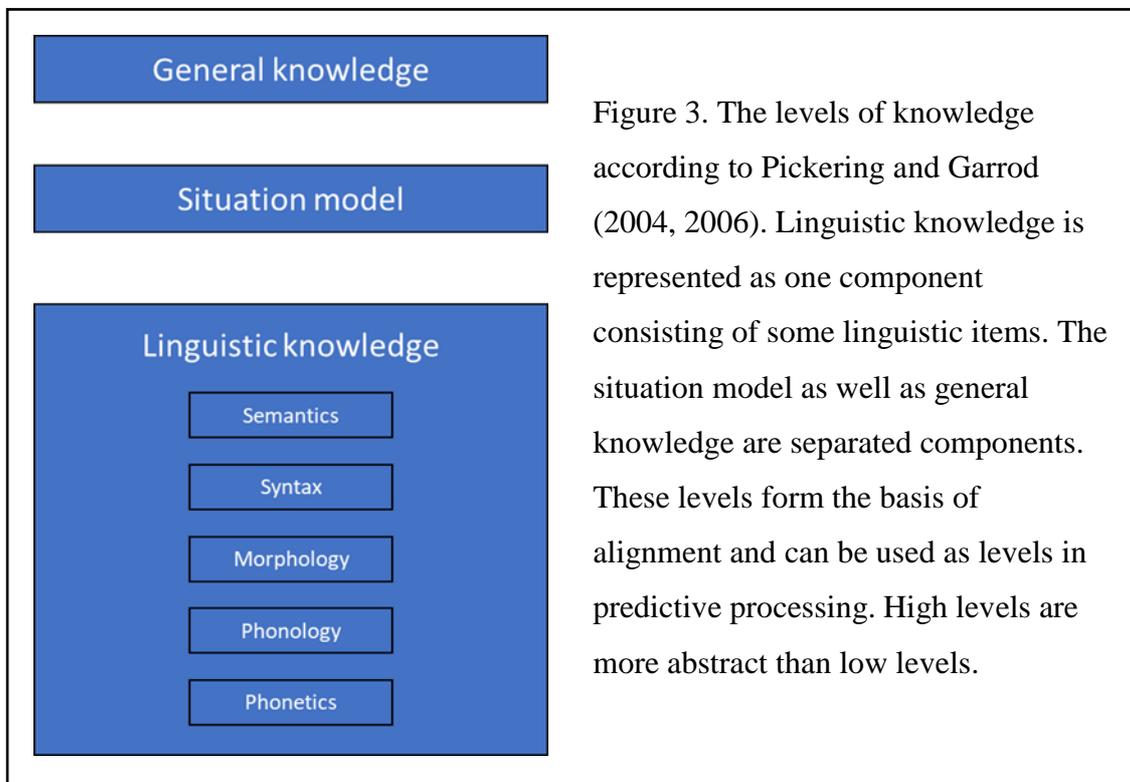
common ground as result of alignment on linguistic knowledge and knowledge about the situation model, as well as on general knowledge (Pickering & Garrod, 2004). This common ground, in turn, leads to utterances that can be understood by both interlocutors, as their situation models are basically the same (Pickering & Garrod, 2004). Thus one is not taking the perspective of the other interlocutor, but one just has (nearly) the same situation model. One can conclude that alignment via priming is the result of already aligned interlocutors. The repair mechanisms, therefore, can be used if interlocutors are not aligned already.

### Capturing interactive alignment with predictive processing

Now that the predictive processing framework, as well as alignment in communication, are explained and defined, it is time to try to combine the two theories as proposed in this thesis. In order to do so, the necessary components of the interactive alignment theory will be modelled with properties or mechanics of the predictive processing framework.

#### Levels of knowledge

First of all, there are three distinct types of knowledge involved in the interactive alignment model: general knowledge, knowledge about the situation model and linguistic knowledge (Pickering & Garrod, 2004). In Figure 2 it is possible to see the different levels. The levels between and including phonetic representation and semantic representation are part of the linguistic knowledge. In predictive processing, there are also levels, which is quite convenient. The levels of alignment can thus be used as levels in the predictive processing framework. A proposal on how to capture this with the predictive processing framework is given in Figure 3. Bear in mind that this is only a representation of the levels and not a Proposal on the whole model, which will be developed in steps during this chapter.



As one can see, general knowledge is the highest level, followed by the situation model and then all linguistic levels. This is in accordance with the theory of Pickering and Garrod (2006) and matches the assumption of predictive processing that higher levels contain more abstract representations and knowledge (Lupyan & Clark, 2015; Pickering & Garrod, 2006).

### **Production and comprehension**

The next step is to connect the just defined levels. In predictive processing, this would be done by predictions and prediction errors. As there cannot be any prediction error without a prediction in the first place, this will be the way to go. Luckily the developers of the interactive alignment theory did add predictions to their model in a later stadium. They assumed that interlocutors do make predictions about what the next utterance will be and compare that prediction to the actual utterance (Pickering & Garrod, 2013). A listener predicts the next utterance of a speaker, while a speaker predicts her own utterance and whether that utterance will be understandable using the own comprehension mechanisms (Garrod & Pickering, 2013; Pickering & Garrod, 2007, 2013). This can be implemented using top-down predictions in predictive processing.

**Production.** Assume person A has a message for person B. Unfortunately there is no distinct description of the message, but in this thesis, the message will be assumed to be underlying all levels and therefore is not a level itself. According to Pickering and Garrod (2013), A then needs to find the right situation model and semantics to utter that message. Hence, A takes into account the possible and most probable utterances and predicts which utterance is most likely to be understood as it should be understood. Thus the content of the hypotheses is, that if A uses certain semantics, then according to A (i.e. his own system), this must be understandable.

In predictive processing, this can be described using predictions. Assuming that there already is a message, which must include a situation model, a user would have the contents of this situation model as hypotheses. The hypotheses would, in turn, cause predictions of how the hypotheses can be best described, i.e. how would the speaker understand these hypotheses. The predictions will then act as hypotheses on the next level, which in this case is syntax and so forth. Figure 4 and 5 give an overview of the levels and how they are combined. Note that every level of Figure 5 can be described using a Bayesian network, such as depicted in Figure 4. Kwisthout et al. (2016) defined a hypothesis as a prior joint probability

distribution over all variables of the hypothesis and predictions as posterior probability distributions over all prediction variables.

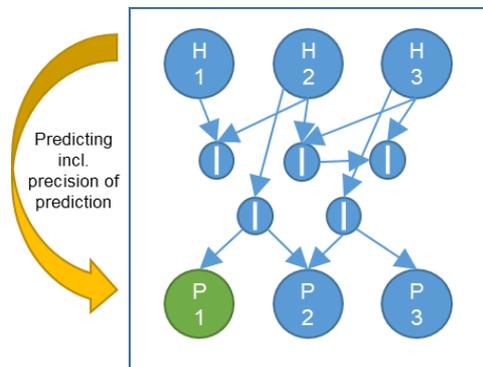


Figure 5. The predictive processing framework as Bayesian network. Adapted from "To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions" by Johan Kwisthout, Harold Bekkering and Iris van Rooij, 2016, *Brain and Cognition* (Vol. 112, pp. 84-91). Copyright 2017 by Elsevier. Adapted with permission.

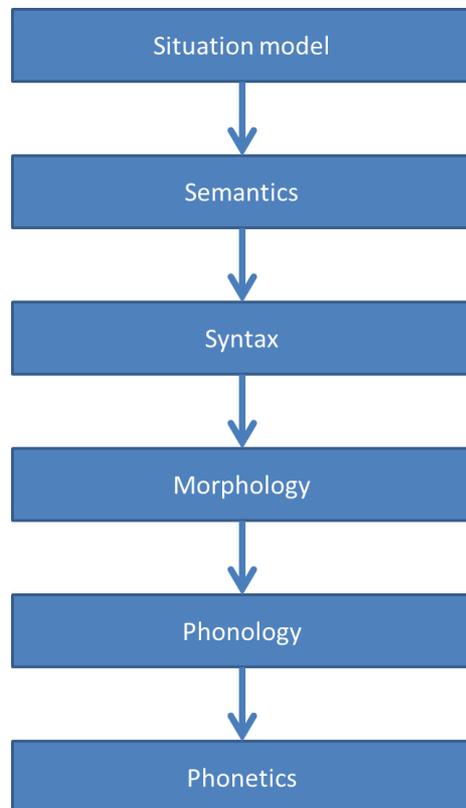


Figure 4. Top-down predictions from situation model to phonetics. Each level contains a causal Bayesian network as depicted in Figure 4. The blue arrows are prediction steps from one level to the next. The message is underlying all levels and activates a situation model. The contents of that situation model act as hypotheses and cause some predictions, which contain information about how the user can make the situation model understandable. This process iterates through all levels until it has reached the phonetical stage. After that, there, of course, will be other instances, such as motor control, which are not being explained in this thesis as the aim is to model higher level cognition.

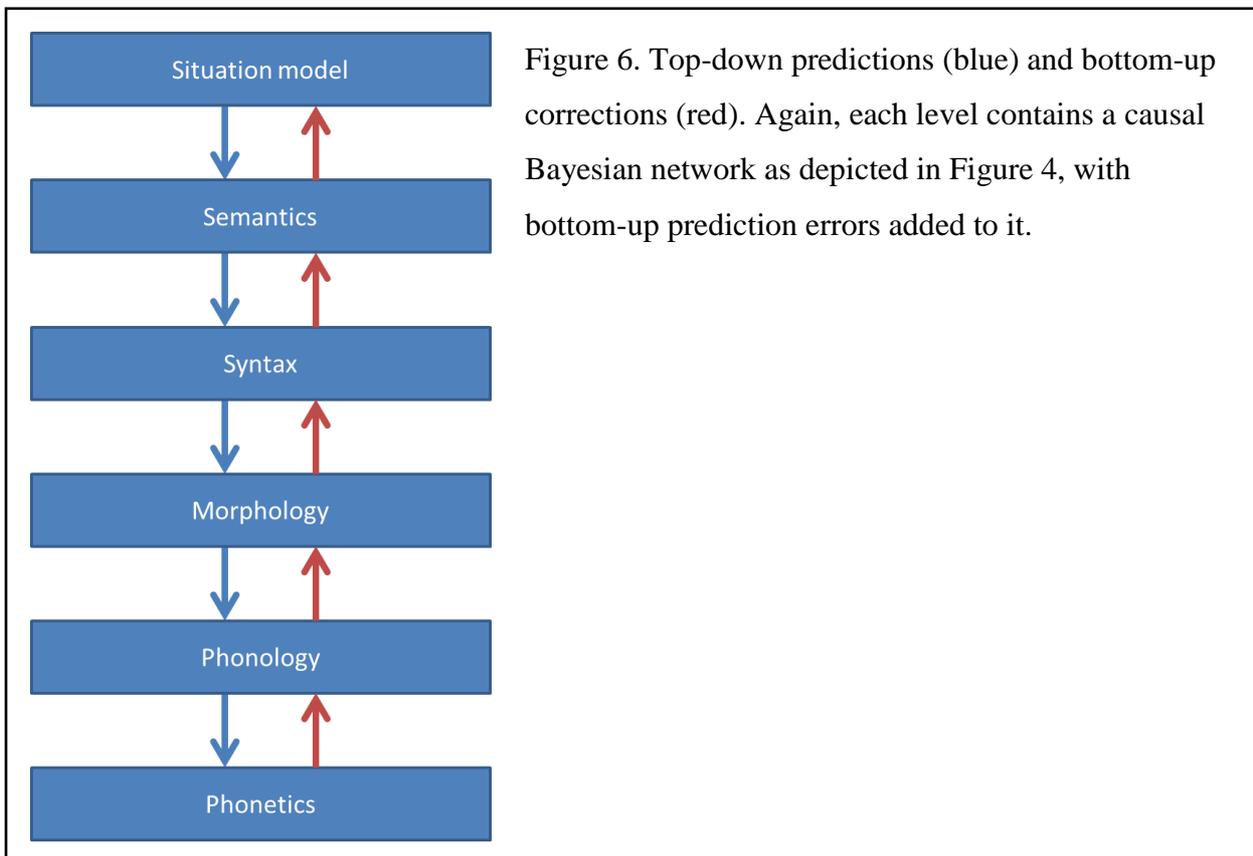
**Comprehension.** The next step is to show how comprehension can be implemented using the predictive processing framework. Garrod and Pickering (2013) introduce the idea that interlocutors predict the utterance of the counterpart using forward models and that this prediction is a prime for their own production system. Listeners are constantly predicting the next utterance by imitating the thus far uttered sentences, understand the underlying message, generate efference copies and then compare those copies to the actual utterance (Garrod & Pickering, 2013).

One could understand that the listener will just use the production system to predict the message, situation model, and the linguistic levels. However, Pickering and Garrod (2013) do point out that a listener only uses the linguistic levels to predict the utterance. They do not specify the situation model nor the message in their account of comprehension. Importantly, Pickering and Garrod (2007) do not say that there are no predictions on higher levels in comprehension. They just do not mention anything about that. It is challenging to capture this phenomenon with the predictive processing framework, as it is a hierarchical framework and depends on the use of predictions from higher levels.

Therefore it is necessary to add some assumptions to the interactive alignment theory of Pickering and Garrod, which will help to model their theory using the predictive processing framework. A proposal is that a listener does make predictions on higher levels. In essence, that would mean that a listener uses the whole production system as defined by Pickering and Garrod to predict the next utterance. This idea is also used in the example of Clark (2015b), where predictions about the world are made, and the following actions then change the real world (A. Clark, 2015b). Assume speaker A is uttering a sentence, for example: “I will get some coffee now”. Listener B will try to predict what A is going to say. Let’s say he predicts that A is going to say: “I will go the bathroom now”. This prediction may be based on B seeing that A is standing up or on the fact that B needs to go to the bathroom. Thus B does actually predict a message and a situation model, which does not need to be right of course if the interlocutors have not aligned yet. Both interlocutors do have active representations on all levels and when listener B hears A say: “I will get some”, he already notes that the prediction cannot be right. Thus the prediction needs to be corrected quite fast. This correction probably happens more often than only one time until the utterance is entirely made and the predictions of B can be fully compared to the utterance of A.

This process can be described using prediction errors in predictive processing. Again assume that B made an incorrect prediction. Then B hears a part of the utterance of A and needs to correct the prediction. This happens on a low level, as the utterance is an observation and is processed bottom-up. In accordance with the theory of Pickering and Garrod (2013), this observation is compared to the lowest level of the model, which is phonetics, and errors are forwarded the model upwards. Comprehension thus first uses top-down predictions and then bottom-up

corrections. When B hears A say “I will get”, B already notices a prediction error which gets forwarded to the next level. Probably this will lead to a prediction error on a semantic level and will also influence the situation model as the utterance “I will get” cannot be used in combination with bathroom and thus B will make new predictions on this and underlying levels.



In Figure 6 prediction errors are added to Figure 4. A listener makes predictions based on what she would say next and based on her own experiences; thus she uses her own conceptual space without taking into account the knowledge of her counterpart. The prediction error is computed as the Kullback-Leibler divergence between the observed and the predicted distribution (Kwisthout et al., 2016). Hence, a listener makes predictions on all levels and observes an utterance. Then he calculates the distance between the prediction and the observation beginning on the phonetic level. The hypothesis on that level may change, and therefore there can be a prediction error on the level above that, because the predictions of that higher level are not in accordance with the hypothesis on the lower level, if the hypothesis is actually updated. This bottom-up error processing can be compared to the percolation process in the theory of Pickering and Garrod (2004). In this context

alignment means that the hypotheses and predictions are updated after hearing the actual utterance; thus the listener aligns his conceptual space to the speaker.

### **Alignment mechanisms**

Note that percolation of alignment is already captured in language comprehension. It is time to integrate priming mechanisms.

**Priming in predictive processing.** According to Pickering and Garrod (2004), priming is the most important mechanism involved in interactive alignment. When a speaker produces an utterance, a representation is activated at all levels (i.e. situational model and linguistic levels) (Pickering & Garrod, 2004). This activation of representations can be denoted as priming on the different levels. The speaker thus primes himself when producing an utterance. For the predictive processing account, that means that predictions are not only based on a message one wants to utter at this moment, but also on previous hypotheses and predictions. This is in accordance with Kwisthout et al. (2016), who described predictive processing using hypotheses and predictions of the past to derive the hypotheses and predictions now. Earlier predictions are being updated most of the time and not fully exchanged.

In comprehension, a listener already has hypotheses and predictions on what the speaker is going to say. When both interlocutors are well aligned, the predictions will probably be quite good (Pickering & Garrod, 2004). Both interlocutors will have the same representations activated on some or all levels, which will lead the interlocutors to use these representations again when speaking or comprehending. In predictive processing this would mean that the predictions of the interlocutors are correct and they use these initial predictions again for the next utterance. Thus, both interlocutors are primed by their own predictions, but also by observations of the conversation. Again it is important to bear in mind that predictions from just a moment before do have an influence on new predictions as they are updated.

Priming, therefore, leads to alignment in the long-run, because interlocutors will, over time, use the same representations for comprehension and production on some or all levels and are therefore aligned. Note that this is mainly the case in already aligned agents or after another alignment mechanism, such as repair, has taken place.

**Repair in predictive processing.** Pickering and Garrod (2006) point out that changes to the situation model are “a normal part of conversation when new

information is introduced”, but that these misalignments are relatively rare in easy dialogue. The repair mechanism is needed when the agents are misaligned, that means when there are indeed prediction errors. As described earlier, these prediction errors also come with a weight called precision. A high precision will lead to the update of one’s predictions or hypothesis, as it means that there is less uncertainty in the prediction error and therefore the error will be weighted heavily. In predictive processing, there are several ways of dealing with prediction errors. These can be used to model the repair mechanisms of the interactive alignment theory and will be described in the following sections.

***Update hypothesis.*** Assume that the speaker did notice that the listener did not understand her message and therefore tries to re-evaluate her utterance. She, therefore, takes her utterance as input and attempts to understand that utterance, thus simulating her own understanding of that utterance (Pickering & Garrod, 2004). If there is a prediction error on a certain level, she can update her hypotheses on these certain levels. This happens if “all uncertainty is irreducible” (Kwisthout et al., 2016), thus when she knows that her utterance was not wrong, but that it was not completely unambiguous. She will make a corrected utterance again and hopefully, the listener will understand. Hence, the hypothesis updating mechanism from predictive processing can be used to model interactive repair using implicit common ground of the interactive alignment theory.

***Update causal model.*** Another possibility is that the listener did understand a part of the utterance, but not everything. Some part of the utterance is entirely new to him, and he needs to update his causal model. For example, if the speaker explains that a star in a mathematical context means that one needs to multiply the values. Assume that the listener already knows how to multiply, but he is not aware that ‘\*’ can mean that one needs to multiply. If he has confidence in the information, but this information is new to him, he can add it to his model. In this case, a new variable is added to the Bayesian network, and the conditional probability distributions are updated. This is another possibility to implement interactive repair using implicit common ground.

***Lowering the level of detail.*** Sometimes one can just reduce the degree of detail of the prediction when there is a prediction error. Talking about the star in math again, the listener could just assume that the speaker is talking about math. However, this will not lead to proper alignment, because the prediction is not wrong

and there will not be a prediction error. That means that there will be no update. Suppose the listener now sees the star in his homework, then he will only know that it has something to do with math, but he has not learned that this star means multiplication. Hence, the interlocutors will not understand each other correctly. Lowering the level of detail does not seem to be a good way to solve prediction errors in communication.

***Gather evidence.*** An agent could also tend to gather new evidence. This update mechanism can be used to describe active repair based on full common ground from the account of Pickering and Garrod (2004). In their active repair mechanism, one possibility is that an agent makes assumptions about the other person to make something more understandable. According to the interactive alignment theory, this does not happen often and is quite slow. Thus, it is a solution when no other alignment mechanism is able to reduce the prediction error. A user will need to add some knowledge about the counterpart to the (predicted) message and the associated representations on the levels, i.e. the situation model and linguistic levels.

Assume that the speaker said: “Today I kicked a ball very hard” and that the listener does not understand that utterance because of some other utterance before which made clear that the speaker is not talking about football. All so far described alignment mechanisms fail to correct this, and the listener does need to make inferences about the speaker. He knows that she plays rugby and just as he thinks about that, he can produce new hypotheses and predictions. Now he does have the same representations active as the speaker, and thus he can understand her. The interlocutors are aligned by using inferences about one another.

***Active inference.*** In predictive processing, active inference means intervening in some situation and thus changing the world. One is not modelling one's own hypotheses or predictions, but one changes the observation actively (Kwisthout et al., 2016). This is quite a radical method to reduce prediction errors and can be used to describe interactive repair making use of full common ground in the interactive alignment theory of Pickering and Garrod (2004). This method will only be used when all the other alignment mechanism failed, as it involves inferences about the other interlocutor, as well as active modelling of situation models, and is therefore quite slow.

Active inference means that the interlocutors are modelling their situation models by “explicit negotiation” or “discussion of the situation models” (Kwisthout et al., 2016; Pickering & Garrod, 2004). Pickering and Garrod (2004) state that such processes are superimposed on the interactive alignment mechanism, and they do not describe how these mechanisms work in detail (Pickering & Garrod, 2004). However, predictive processing could be able to capture these mechanisms too. Assume that one is talking about a friend called ‘James’. Both interlocutors know two persons with that name. The listener is not able to understand which James is meant, so he asks: “Do you mean James with the blond hair?”. The listener already made the inference that the speaker knows a James with blond hair and actively talks about his situation model. The speaker could then reply: “No, I mean James with the cat”. In this example, both interlocutors have actively made inferences about the other one, and they talked about their situation model. The speaker tried to model the situation model of the listener so that the listener will think of a James with a cat.

In predictive processing, this would mean that there is a prediction error that cannot be solved using the other methods and therefore the speaker needs to change the world actively and thus model the situation model, i.e. the hypotheses or predictions, of the listener.

**Routines.** Pickering and Garrod (2004) used routines to describe fixed expressions with a high frequency. An example of such a routine is the expression: “How are you?”. There are two possibilities to represent such routines in the thus far developed model. The first one is to summarise all linguistic levels for these expressions. That means, the user needs to have a message and a situation model, but the linguistic levels are directly linked to that situation model. However, this is not in accordance with the predictive processing framework as one would just skip some levels of processing. If one needs to go through all levels in one case, then one needs to go through all the same levels in another situation too. Levels cannot arbitrarily be changed or summarised in predictive processing.

Hence, routines need to be described in another way. This could be done using probability distributions in predictive processing. Pickering and Garrod (2004) define a routine as an expression with a high frequency. In predictive processing, this would need to mean that the probability of a certain chain of representations (on linguistic levels) is higher than others. As a consequence, after a

situation model is made up, one is more likely to use that expression again in combination with the situation model. However, there is a problem in modelling such routines with the predictive processing framework. As described in Figure 1, in predictive processing only prediction errors are being used to update probabilities of the probability distributions of the levels. When an agent made correct predictions, there will be no changes to the hypothesis, predictions or probabilities. The probability of such a routine would only be modified if there were a prediction error. This problem can be addressed by adding another assumption to the predictive processing framework. It should be considered, that a right prediction will lead to a change of the probabilities of that chain of predictions. Suppose the overall prediction of the utterance “How are you” has been made and turns out to be a right prediction. In today’s theory of predictive processing, this would not lead to any changes. Why are those accurate predictions not used to increase the probability of it? When a prediction is being confirmed, and therefore another experience has been made, the probability of that prediction should be increased.

Pickering and Garrod (2004) did point out, that some routines are only active during one conversation and are not used again after that conversation. This aspect of routines cannot be explained with conditional probability distributions as used for long-term routines, but as explained earlier, predictions from the moment before do have an influence on predictions now as previous predictions are only updated. Such short-term routines can be explained by this fact. The predictions of the moment just before, are a basis for the predictions now, but after the conversation has ended, they do not influence new predictions.

Note that routines in predictive processing are not exactly what Pickering and Garrod (2004) had in mind. They defined routines as stored representations of certain expressions, whereas in the context of predictive processing, one needs to characterise routines as high probabilities in the probability distributions through the levels. By describing routines with these probability distributions, the most important feature, which is fast usability of these expressions, gets lost. Agents must compute the probability distributions again, and the use of routines does not accelerate anything.

**Monitor.** The interactive alignment theory also contains self-monitoring. Agents are continuously monitoring themselves on each level of the model. Users ask themselves if the output of that level is in accordance with the implicit common ground (Pickering & Garrod, 2004). More detailed, Pickering and Garrod (2004) describe that a speaker is listening to himself, not only in the sense that he listens to the actual sounds, but also internally and automatically on every level. In predictive processing, monitoring could just mean that a speaker, while producing an utterance, is listening to himself and checking internally on every level, if the output of the levels is in accordance with the implicit common ground, i.e. with what the user wants to say and how he wants to say it. Thus a speaker is not only producing but also comprehending his own utterance at the same time and in the same system. When a prediction has been made, the output is directly used as input again, and an error in this process will lead to alignment through a repair mechanism. Hence, the speaker will change the utterance while speaking or before he starts to speak.

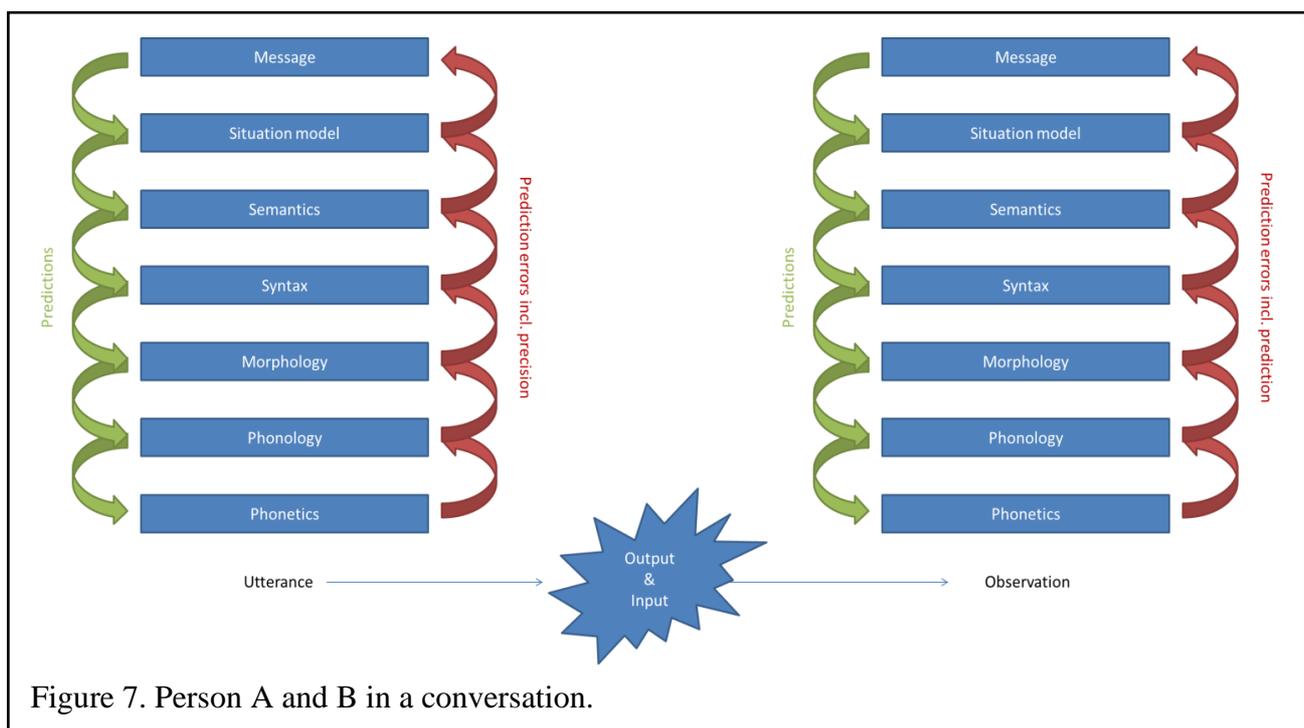


Figure 7. Person A and B in a conversation.

Figure 7 gives an overview of how a discussion between two interlocutors would work. Each level can be represented with a causal Bayesian network as shown in Figure 5, with bottom-up prediction errors added to it. Table 1 again gives an overview on how the interactive alignment theory has been modelled with the predictive processing framework in this thesis.

Table 1. Interactive alignment and predictive processing combined.

Interactive Alignment	Predictive Processing	Combination
Levels of knowledge	Hierarchical levels: abstract (top) to concrete (bottom)	From abstract to concrete: General knowledge - Situation model - Linguistic knowledge
Language production	Top-down predictions	Top-down predictions are being uttered.
Language comprehension	Top-down predictions & Bottom-up prediction errors	Predict next utterance and compare prediction to observation. Errors flow bottom-up through the levels, leading to updates.
Percolation	Bottom-up prediction errors	Errors percolate bottom-up through levels. Alignment on a lower level leads to alignment on a higher level.
Priming	Time-dependency & Predictions	User tends to use the same representations/hypotheses again. By predicting, comprehending, or producing, some representations are active and are used again in the prediction/comprehension/production.
Repair mechanisms	Handling prediction errors	Misalignment has happened, and the beliefs need to be revised, or other updates need to be made.
Interactive repair using implicit common ground	Hypothesis updating	Notice misalignment and re-evaluate own utterance. Update the hypothesis, so that the utterance is unambiguous and utter again.
Interactive Repair using implicit common ground	Update causal model	When new information is introduced, a new variable needs to be added to the probability distribution of particular levels.
Interactive repair using full common ground	Gather evidence	Inferences about the counterpart to understand the utterance.
Interactive repair using full common ground	Active inference	Inferences about counterpart and model counterparts' situation model by explicit negotiation.
Routines	Probabilities	Higher probability of a certain expression.
Routinization	Time-dependency	Update hypotheses and use same representations.
Monitor	Output as Input	Output is immediately used again as input.

## Discussion

The aim of this thesis was to capture the interactive alignment phenomena in communication described by Pickering and Garrod using the predictive processing framework. In order to do that, it was necessary first to define alignment and predictive processing, and after that, the theories have been brought together. Alignment is reached when representations on some or all levels are the same in both interlocutors (Pickering & Garrod, 2004). By using the predictive processing framework to capture that, it has been shown that the mechanisms of the interactive alignment theory can be modelled with the mechanisms in predictive processing. When the levels of Pickering and Garrod (2004) are described by Bayesian networks with corresponding probability distributions as in the notation of Kwisthout et al. (2016), one can use hypothesis updating, updating of the causal model, evidence gathering and active inference as alignment mechanisms. These mechanisms lead to aligned representations on the different levels and can, therefore, describe alignment mechanisms defined by Pickering and Garrod (2004). Percolation between the levels is inherent in predictive processing and has been described as errors percolating through the levels in a bottom-up direction. However, according to the interactive alignment theory, there are only repair mechanisms when misalignment has happened already (Pickering & Garrod, 2004).

The main alignment mechanism is priming, and the underlying assumption is, that previous predictions and the thus far perceived dialogue do have an influence on new predictions and utterances. Pickering and Garrod (2004) highlight the importance of priming, i.e. the just heard and predicted utterances have activated some representations and will have an influence on future utterances and predictions. Kwisthout et al. (2016) show that earlier predictions do have an impact on current predictions in the framework as predictions are being updated and not just thrown away. It has been demonstrated that priming can have an important role when talking about alignment, but other mechanisms, such as repair are essential to the theory. Realistically speaking, interlocutors are not already aligned, and therefore some updating with repair mechanisms is needed. This is a major flaw in the interactive alignment theory. However, after some representations are aligned, great parts of the conversation can be described using priming. When the probabilities or hypotheses are being updated, that would also have quite an influence on other dialogues in the future, but for priming, this effect should be only

short-term in most cases. The just activated representation will change ones next utterance, but it will not necessarily change one's next conversation with another interlocutor. Only when an interlocutor is primed relatively often, the model will undergo long-term changes.

Furthermore, Pickering and Garrod (2004) described, that in priming the same or nearly the same representations get activated in both interlocutors. This can be the case if both interlocutors do know each other already and they do already have an implicit common ground, but strangers will not have the same representations that often. Hence, as found in this thesis, repair mechanisms would be needed first until some implicit common ground has been developed. After that, priming can have some influence and will probably be an important mechanism in the interactive alignment theory. The importance of priming depends on the definition of dialogue. Pickering and Garrod (2004) pointed out that their theory is restricted to easy dialogue only. Still, it is not obvious what that means. In this thesis, it has been found that priming can have an influence after two agents are already aligned. Consider two good friends in a casual dialogue or a dialogue between a parent and a child, where no new information is being provided. This could be examples of easy dialogue, where the repair mechanisms are not needed that much and priming will be the major mechanism. On the other hand, consider the time before two interlocutors were friends or a teacher – student situation, where there is quite a lot of new information. In such situations, repair mechanisms are highly needed, and priming will not be sufficient to achieve mutual understanding in both agents.

One of the strategies in the predictive processing framework described in this thesis is to lower the level of detail of one's predictions. As described earlier this cannot be a sufficient alignment mechanism. When the degree of detail is lowered, that means that hypotheses are more generic, which will not lead to alignment in the sense of similar representations on the levels. This method could be used to describe a process in which the interlocutors do not understand each other, i.e. do not have the same representations, but they do not actually bother to understand all the details.

The predictive processing is not able to capture routines as described in the interactive alignment theory. It has been shown that such routines can be captured, if another assumption is added to the predictive processing framework. Correct predictions should also have an influence on the probability distributions on the

different levels. If another experience has been made and predicted, that experience should strengthen the probability of the prediction. By adding that assumption to the predictive processing framework, it is possible to model expressions with a higher frequency. However, this is still not fully in accordance with the definition of routines of Pickering and Garrod. The advantage of such routines is that processing is much faster compared to the processing of other expressions. This is an advantage that cannot be described by predictive processing, as it is a hierarchical framework and the probabilities must be calculated for every expression.

The last described mechanism in this thesis was monitoring. It was defined as a speaker who is monitoring his own outputs per level when producing an utterance. One could question the conformity with the predictive processing framework. In predictive processing prediction errors are processed in a bottom-up fashion. Using the description of Pickering and Garrod (2004) however, will not lead to such bottom-up processing, as they state that agents are monitoring every level and errors can result in a change of the utterance. This infringes core beliefs of predictive processing: bottom-up processing of prediction errors and the hierarchical nature thereof. Here it has been proposed to use the output of the forward production mechanism as comprehension-input to the same system. This would mean that the predictions are being compared to the output and can lead to prediction errors, which, in turn, can lead to an internal update via the described mechanisms. However, this is not exactly the monitoring mechanism as described in the interactive alignment theory, as this cannot be done per level individually but has to be done bottom-up through all levels. It would be exciting to see how monitoring in the interactive alignment can be changed such that it can be modelled with the predictive processing framework.

To capture interactive alignment phenomena in communication, the term message has been used. Unfortunately, both, the interactive alignment theory and the predictive processing framework do not further describe where these messages come from in the first place. In predictive processing, there is an ongoing loop of predictions, but there is no description of how the first thought is being derived. Pickering and Garrod (2004) assume that an interlocutor has a message, no matter how that message is derived. Future research should show how interlocutors derive a certain message. This will help to understand human cognition on a higher level. Perhaps it is possible to describe the derivation of such messages with predictive

processing; thus one could add more levels to the developed model. Please be aware that the question of such a zero point and the origin of predictions, messages and thoughts can be seen as a knockout argument for most or perhaps every theory at the moment.

As the predictive processing framework should be able to capture all cognitive processes, it will be very interesting to see if it is possible to capture communication in its completeness. This thesis did not take into account non-verbal communication, which plays a crucial role in communication (Bavelas & Chovil, 2006). In order to do that, the predictive processing framework must be able to work with multimodal inputs (Galati & Brennan, 2014). Hopefully, the content of this thesis can act as a starting point and inspiration for future research on how to model alignment in communication including gestures and facial expressions with the predictive processing framework.

This thesis has shown that it is possible to capture general alignment mechanisms described by Pickering and Garrod with the predictive processing framework, but also that there are some challenges which should be tackled in future research. The interactive alignment theory is meant to describe easy dialogue only, but with the predictive processing framework, it is possible to describe more complicated processes, such as repair mechanisms using full common ground. It was also possible to describe the repair mechanisms in more detail, as both can be described with respectively two update mechanisms of the predictive processing framework. The levels of knowledge, priming, all repair mechanisms of the interactive alignment theory and complex monitoring have been modelled with the predictive processing framework. It has been argued that priming is essential in already aligned interlocutors, but that repair mechanisms are needed in order to describe dialogue where new information is being introduced. Hence, with its explanatory power and neurophysiological evidence, the predictive processing framework can broaden and strengthen the interactive alignment theory, while also pinpointing to the flaws of it.

### Acknowledgements

I would like to thank my supervisor Dr Mark Blokpoel for his ongoing support, motivation, inspiration and patience. I would also like to acknowledge Dr Johan Kwisthout for organising a research seminar on predictive processing and allowing me to join it.

## References

- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1-22.
- Bavelas, J. B., & Chovil, N. (2006). Nonverbal and Verbal Communication: Hand Gestures and Facial Displays as Part of Language Use in Face-to-face Dialogue.
- Bell, A. (1984). Language style as audience design. *Language in society*, 13(02), 145-204.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3), 355-387.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13-B25.
- Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24(6), 489-506.
- Brown, H., & Friston, K. J. (2012). Free-energy and illusions: the cornsweet effect. *Frontiers in psychology*, 3, 43.
- Chandler, D. (1994). The transmission model of communication.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181-204.
- Clark, A. (2015a). Radical predictive processing. *The Southern Journal of Philosophy*, 53(S1), 3-27.
- Clark, A. (2015b). *Surfing uncertainty: Prediction, action, and the embodied mind*: Oxford University Press.
- Clark, H. H. (1996). Communities, commonalities, and communication. *Rethinking linguistic relativity*, 17, 324-355.
- Communication. (2017). Retrieved from <http://www.oxforddictionaries.com/definition/communication>
- Doya, K. (2007). *Bayesian brain: Probabilistic approaches to neural coding*: MIT press.
- Eckert, P., & Rickford, J. R. (2001). *Style and sociolinguistic variation*: Cambridge University Press.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1456), 815-836.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138.
- Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *cortex*, 68, 129-143.
- Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, 15(17), R644-R645.
- Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, 29(4), 435-451.
- Garcia, D. L. (1996). Global Communications: Opportunities for Trade and Aid. *SAIS Review*, 16(1), 35-66.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181-218.
- Garrod, S., & Pickering, M. J. (2013). Interactive alignment and prediction in dialogue. *Alignment in Communication: Towards a new theory of communication*, 6.

- Giles, H., Coupland, N., & Coupland, I. (1991). 1. Accommodation theory: Communication, context, and. *Contexts of accommodation: Developments in applied sociolinguistics, 1*.
- Harkness, D. L., & Keshava, A. (2017). Moving from the What to the How and Where – Bayesian Models and Predictive Processing. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.
- Hohwy, J. (2013). *The predictive mind*: Oxford University Press.
- Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 580-593.
- Johnson-Laird, P. N. (1980). Mental Models in Cognitive Science. *Cognitive Science, 4*(1), 71-115. doi:10.1207/s15516709cog0401\_4
- Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Phil. Trans. R. Soc. B, 370*(1668), 20140169.
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation, 32*, 385-418.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review, 94*(2), 211.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences, 27*(12), 712-719.
- Kwisthout, J., Bekkering, H., & van Rooij, I. (2016). To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and cognition*.
- Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological review, 94*(4), 412.
- Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. In *Current Directions in Psychological Science* (Vol. 24, pp. 279-284).
- Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol, 7*(1), e1001048.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences, 27*(02), 169-190.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language & Computation, 4*(2), 203-228.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in cognitive sciences, 11*(3), 105-110.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329-347.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language, 696-735*.
- Seth, A. K. (2014). The cybernetic Bayesian brain. In *Open Mind: Open MIND*. Frankfurt am Main: MIND Group.
- Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3-55.
- Stevens, S. S. (1950). Introduction: A definition of communication. *The Journal of the Acoustical Society of America, 22*(6), 689-690.

- Thornton, C. (2016a). Predictive processing is Turing complete: A new view of computation in the brain.
- Thornton, C. (2016b). Predictive processing simplified: The infotropic machine. *Brain and cognition*.
- Tomasello, M. (2009). *The cultural origins of human cognition*: Harvard University Press.
- Wachsmuth, I., De Ruiter, J., Jaecks, P., & Kopp, S. (2013). *Alignment in Communication: Towards a new theory of communication* (Vol. 6): John Benjamins Publishing.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162.