The wisest or the nicest?

The influence of speaker certainty and affect in word learning

Master thesis Myrthe Bergstra, Utrecht University   Supervisors: Peter Coopmans and Hannah De Mulder

ABSTRACT – This study aimed to find out whether speaker certainty of the interlocutor and affect towards the interlocutor could guide children in the word learning process. In three experiments with 4- to 6-year-old children, two handpuppets labeled novel objects. It was found that children tend to learn new words from a certain speaker rather than from an uncertain speaker, but they had no significant preference for learning from a kind speaker. When both cues were combined, it was found that children prefer an unkind and certain speaker over a kind and uncertain speaker, showing that speaker certainty leads to stronger learning preferences than affect. However, kind and certain speakers were preferred most, indicating that affect has a small additive effect. A posttest in which children had to recall which object belonged to which label showed that many of the new links between labels and objects were enduring.

Introduction

It appears that we build our vocabulary by learning word by word. In doing this, we depend on the people around us, who label objects in our presence. Recently, it has been found that children not only make use of object-focused strategies, but also of cues from their interlocutor during this labeling process. These cues might indicate for example which speaker is reliable to learn from and which speaker is less reliable. Several studies have shown that these cues lead to preferences in children's learning. For example, when a speaker who labels familiar objects correctly is contrasted with a speaker who makes mistakes in labeling familiar objects, children are more likely to accept labels for novel objects from the speaker who was previously accurate in object labeling than from the speaker who was previously inaccurate (see for example Koenig, Clément & Harris, 2004). Children also tend to learn from a speaker who is certain of his object labeling, rather than from one who is uncertain (Bergstra, De Mulder & Coopmans, 2013). Moreover, more general reliability features of the interlocutor, such as their intentions, play a role as well. Doebel and Koenig (2013) argued that having good intentions might lead to general reliability, and therefore to reliability in word learning situations. They showed that when a moral speaker and a neutral speaker labeled objects, children were more likely to learn a new word from the moral speaker than from the neutral speaker. When a neutral speaker and an immoral speaker were contrasted, they were more likely to learn a new word from the neutral speaker.

Selective (word) learning strategies

Recent research has shown that not only traditionally investigated object-focused strategies, such as assumptions on possible referents of words (cf. Markman, 1994), may guide the child in word learning, but that the interlocutor - the person who is uttering the word - may also provide cues. The child's focus on the interlocutor is already apparent at a very young age, as is shown in joint attention studies. For example, Moore et al. (1999) showed that 24-month-old children were guided by gaze direction rather than salience in determining the referent of a novel label. When two objects were present, the children chose the object that the adult was looking at as the referent of the novel label, even though another object was more salient (i.e. it was lit up). Interlocutors may thus be important in guiding the child, but certain features of an interlocutor can also cause a child not to learn from that individual at all. Recent research shows that children do not simply adopt everything from everyone; they are rather selective in their learning (Koenig & Sabbagh, 2013). Children both consider the information that speakers provide them with as well as the general nature of the speaker (Koenig & Sabbagh, 2013). These cues might lead to a preference for learning from one particular speaker over learning from another speaker. This phenomenon will henceforth be referred to as *selective learning*.

A typical paradigm used in selective learning studies in general presents children with two informants. In most cases, the children first witness something that gives information about these informants. For example, the experimenter gives information by telling the child that one informant is a liar. Or the informants could display certain behavior that illustrates their difference in reliability, such as answering questions, in which one informant provides correct answers and the other informant provides incorrect answers. The children's preference for a certain informant is measured for example by tracking whose advice the children follow, who they think is more helpful, or from whom they learn something (Mills, 2013). In word learning studies, this general paradigm is often administered, and trials typically consist of two informants labeling novel objects. For example, two informants name two different objects using the same label (see for example Birch, Vauthier & Bloom, 2008); informant 1 calls object X a "ferber", and informant 2 calls object Y a "ferber". The child is then asked which object is the "ferber". Whether children choose object X or Y shows which informant they found reliable. In most studies, the difference between the informants becomes clear when introducing the informants, but in some studies, the difference between the speakers becomes clear in the trials themselves; for example when they use different formulations in

their labeling. Two cues that have been investigated in these kinds of setting and have been shown to lead to selective learning are knowledgeability and prior accuracy of speakers.

Knowledgeable and accurate speakers

One of the first to investigate selective learning were Sabbagh and Baldwin (2001). Although their design is slightly different than those of others, because they used only one informant, the study provided interesting results by showing that children are more willing to accept label-object pairing from a speaker when he is knowledgeable than when he is ignorant. To test this, 3- and 4-year old children were presented with a single speaker, who was knowledgeable in one condition and ignorant in another condition. In both conditions, the experimenter first explained to the child that the playroom was shared with a friend named Birdie, and that Birdie sometimes asked the experimenter to hand her a toy. Birdie then asked the experimenter to hand her a toy, for example, her "blicket". In the knowledgeable condition, the experimenter said: "You know, I'd like to help my friend Birdie, and I know just which one is her blicket. It's this one.", and then touched the target toy. In the ignorant condition, however, the speaker said: "You know, I'd like to help my friend Birdie, but I don't know what a blicket is. Hmmm. Maybe, it's this one.", and he then touched the target toy (Sabbagh & Baldwin 2001: 1057). Afterwards, children had to name pictures, among which the target toy was present. Children were more likely to call the target toy a "blicket" in the knowledgeable condition than in the ignorant condition. Children were also asked to get the "blicket" from a collection of six toys and put it in a box. They were more likely to get the right toy in the knowledgeable condition than in the ignorant condition. In short, children were more likely to accept the label-object pairing when the speaker was knowledgeable than when the speaker was ignorant.

Koenig, Clément and Harris (2004) investigated whether prior accuracy of an informant could lead to selective learning by children. They presented 3- and 4-year-olds with a video of two informants who labeled objects. In the familiarization phase, the informants labeled familiar items. One of them was accurate in his labeling (e.g. saying "That's a ball." when a ball was presented) and one of them was inaccurate (e.g. saying "That's a shoe." when a ball was presented). After this, the children were explicitly asked to identify either the person who was accurate or the person who was inaccurate. In the test trials, the informants labeled unfamiliar objects with novel (pseudo word) labels. For example, informant 1 called an object a "modi", while informant 2 called the same object a "toma". The children were then asked what the

object was called, "modi" or "toma". Afterwards, they were again asked to identify the accurate or inaccurate informant. Koenig et al. (2004) found a relation between the children's performance on the test items and the identification of accuracy: the children who correctly identified the accurate or inaccurate speaker performed above chance level at the test trials (i.e., they used the label of the accurate informant more than at chance level), while the children who failed to identify the accurate or inaccurate speaker scored at chance level at the test trials. This shows that children preferred to learn new words from a speaker who was previously accurate at object labeling over learning new words from a speaker who was previously inaccurate at object labeling. Birch et al. (2008) replicated these results by means of a similar paradigm, in which two informants first labeled familiar objects in a familiarization phase and then novel objects in the testing phase. However, they did not ask the children to explicitly identify the accurate or inaccurate speaker. Their findings show that children were more likely to accept the label-object linking from the accurate speaker than from the inaccurate speaker. Since the children were not asked about the speaker's accuracy, Birch et al. (2008) point out that children *spontaneously* keep track of an informant's history in accuracy.

Although these experiments show that knowledgeability and accuracy lead to selective trust (i.e., the children accept a label-object pairing rather from one speaker than from another), we cannot conclude on the basis of these experiments that the children indeed *learn* the new words. It is possible that the children just decide to "stick with the accurate person", without storing the new word in their lexicon (Birch et al., 2008). In order to clarify this issue, the experiment of Birch et al. (2008) had a second condition: the contrast condition. In this condition, the children were not asked for the "ferber", the label that had been used, but for the "modi". On the basis of the mutual exclusivity constraint (cf. Markman, 1994), the children could infer that "modi" was the real name for the object that the inaccurate person had called a "ferber". It turned out that in this condition, the children picked the object of the accurate speaker - so the real "ferber" - significantly below chance, in other words, they had learned the label-object mapping of the accurate speaker and inferred that the other label must be connected to the other object. This shows that the children did not simply "stick with the accurate person" in their answers, but really displayed selective learning behavior. The concern that children might not really learn the new words (Birch et al., 2008) will also be addressed in the present study by investigating whether the label–object link is an enduring link.

The above mentioned experiments showed the effect of interlocutor cues like prior accuracy on word learning. However, Doebel & Koenig (2013) proposed that more general reliability features could also influence the child in deciding from who to learn. Interlocutor cues in word learning then do not have to be limited to cues directly linked to word labeling and could be as general as kindness of a speaker; whether the child will be more likely to learn from a speaker that is kind (and the child therefore likes this speaker better than another speaker) which will be investigated in the present study. It has already been shown that within the domain of language, reliability is considered by the children on a general level: prior accuracy of a speaker in one domain of language can be generalized to the word learning domain. This was shown by Sobel & Macris (2013), who investigated whether prior accuracy in the syntactic domain caused selective learning behavior in word learning by 4-year-olds, i.e., whether prior accuracy in another domain of language was generalized to word learning. They presented children with two speakers. One of them spoke correctly, the other one gave accurate lexical information but erred on syntax and violated subject-verb agreement (e.g. by saying "That are a ball" (Sobel & Macris, 2013: 524)). The question was whether children would infer from this unreliability in the domain of syntax that the speaker would be unreliable in other domains of language as well, and would therefore not be trusted enough to learn from him. In the training phase of the experiment, two speakers labeled familiar objects. One of them used incorrect syntax. In the test phase, both speakers labeled a novel object by means of one single novel word (no syntax was used). The child was asked what she thought the object was called. Sobel & Macris (2013) found that the older half of the 4-year-olds (but not the younger children) displayed selective learning behavior; they significantly more often chose the word label of the speaker who used accurate syntax. Prior accuracy in another domain of language (namely syntax) was thus indeed generalized to the domain of word learning.

General reliability features
Recent research follows the prediction of Doebel & Koenig (2013) that general reliability features can lead to learning preferences: children also display selective learning on the basis of general speaker features which seem not only not to be related specifically to word learning, but also seem not to be related specifically to language. One of these features is age. When presented with two speakers, an adult and a child who provide accurate, but different labels (e.g. calling a specific type of shoe either *shoe* or *sneaker*), 3- and 4-year-old children chose the label of the adult significantly above chance (Jaswal & Neely, 2006). Although at

first there might not seem to be a reason why adults should be more reliable at word labeling than children, one could imagine that children are in general more willing to accept information from adults than from other children. This preference for adults is not completely persistent, however, as is shown when children are presented with an inaccurate adult speaker and an accurate child speaker. In this scenario, children chose the novel label of the child speaker significantly more often, indicating that prior accuracy of a speaker is a more salient feature in the word learning process than a speaker's age (Jaswal & Neely, 2006).

Not only age, but also native accent is a feature that might make speakers more reliable to children and therefore guides them in their word learning process. Corriveau, Kinzler and Harris (2013) presented 3-, 4- and 5-year-old children with two speakers. In the familiarization phase of their experiment, the children heard the speakers speak English, one of them with a native accent and the other one with a foreign (Spanish) accent. Next, the children were presented with a novel object and asked which speaker they wanted to ask what the object was called. Both speakers then named the object (each with a different novel label), and the child was asked what she thought the object was called. Children of all three age groups preferred the native accented speaker, both in the "who do you want to ask"-questions and in the "what is the object called"-question. This experiment thus shows that native accent leads to selective learning in word learning, confirming that the features of a speaker that guide children do not need to be directly linked to the domain of word learning, but can also be more general reliability features. However, similar to the preference for adults, the preference for native accent is not completely persistent: it can be overruled by the effect of accuracy (Corriveau et al., 2013). After the speakers had named the novel objects, they named some familiar objects. One speaker was completely accurate, the other one was completely inaccurate. Then, the speakers again named novel objects and the children were asked again who they wanted to ask what the objects were called and what the objects were called. It was found that when the native-accented speaker was also the accurate speaker, all three age groups performed above chance in following the labeling of the native speaker. However, when the native-accented speaker was the inaccurate speaker, the 3-year-olds performed at chance level. The 4- and 5-year-olds then chose the object label of the accurate, but foreign-accented speaker significantly more often. This shows that for 4- and 5-year-olds accuracy of a speaker is a more salient feature in the word learning process than native accent, but that 3-year-old children rely on native-accented speakers rather than accurate speakers.

There are thus several features that could make an interlocutor seem reliable to a child, which are not all related directly to word learning, but seem to be general indicators of whether a speaker is trustworthy to learn from. Another factor that is not related to word learning, but is related to the novel objects of which children learn the labels, is the informants' expertise about the objects. Sobel and Corriveau (2010) conducted an experiment in which novel objects could make a machine turn on a red light or a green light. In the training phase, two informants told the children what the objects would do. One informant was a "green expert", and predicted correctly which objects would make the light turn green, but said "I don't know" for the other objects. The other informant was a "red expert" and predicted correctly which objects would make the light turn red (and said "I don't know" for the other objects). In the test phase, objects were placed on the machine but the informants did not say anything about what would happen. The machine light then turned either red or green. Then the experimenter said that the object had a special name and both informants labeled the object (with different names). The child was then asked what she thought the object was called. Sobel and Corriveau (2010) found that 4-year-old children chose the label of the expert informant significantly above chance level. These data show that children consider an informant's expertise on an object during word learning.

Generalizability of the findings

Lucas, Lewis, Pala, Wong and Berridge (2013) have pointed at the fact that most studies concerning selective learning have been conducted in English-speaking countries. They conducted a study in which they compared the performance of Turkish, Chinese and English children, to find out if different cultures and languages could lead to different performance of the children. An interesting difference between these populations is that Turkish children speak a language in which evidential markers are used (as opposed to Chinese and English). Evidential markers are markers that indicate the source of someone's knowledge, for example if one knows about an event because one has seen it happen, or because one has been told about the event (Lucas et al., 2013). Although the source of one's knowledge does not necessarily say something about a speaker's certainty about what he is saying, in general knowledge based on one's own perception is viewed as more reliable than more indirect knowledge (Lucas et al., 2013). Speakers of Turkish are forced to think about the sources of their knowledge when expressing this knowledge. When one thinks about a source of certain knowledge, one possibly also thinks about the reliability of the knowledge. This has led to the question whether growing up with a language with evidential markers might facilitate

children's awareness of differences in reliability between speakers. In other words, are Turkish children better at selective learning than Chinese and English children? Lucas et al. (2013) presented children with a selective trust paradigm in which two informants were labeling objects. The informants differed in prior accuracy on familiar objects. It was found that Turkish children were, in general, over 11 times more likely to follow the accurate informant than the other children. Moreover, the Turkish children formed the only group that were flexible in their trust; they trusted an informant that was a toy expert on toy items but not on food items, and an informant that was a food expert on food items but not on toy items. This shows that Turkish children's understanding of reliability is also more sophisticated than that of English and Chinese children (Lucas et al., 2013). However, Sobel and Corriveau (2010, study mentioned above) showed that English children are also capable of considering an informant's expertise on certain objects but not other objects. Perhaps this difference in results is due to the age of the participants; Lucas et al. (2013) included 3- and 4-year-olds, while Sobel and Corriveau (2010) included 4-year-olds only, who may have been more developed cognitively. Both English and Turkish children might thus be able to consider expertise of the informant, but Turkish children seem to be able to do this at an earlier age. The present study will focus on Dutch children. Since the Dutch culture is quite similar to English culture (western world), and Dutch contains no evidential markers, the expectation is that Dutch children will be able to consider interlocutor features in word learning at age 4 and 5, similar to English speaking children.

Different native languages could thus lead to different performance in selective learning studies. It is, however, still not quite clear which (cognitive) demands underlie the ability to consider interlocutor features in word learning. One factor that has been suggested to be relevant is Theory of Mind (cf. Lucas et al., 2013, Bergstra et al., 2013, Birch et al., 2008).

Speaker certainty and Theory of Mind
Theory of Mind, in short the ability to understand that someone else's beliefs may differ from one's own, usually develops around the age of four. It has been suggested that Theory of Mind (ToM) is relevant in word learning, since consideration of other people's mental states is relevant for determining someone's reliability (for example, to know if someone is lying, one needs to know what the other person knows).

Theory of Mind has been shown to correlate with speaker certainty expressed by mental state verbs (Moore, Pure and Furrow, 1990). Speaker certainty is the amount of certainty that a speaker has towards what he says. A speaker can be very sure about something and say, for example "I *know* my keys are in my bag", or be unsure and say, for example "I *think* my keys are in my bag." Words like *know* and *think* are called mental state verbs, because they indicate one's mental state; the attitude that one has towards a belief. Moore et al. (1990) showed that children start to understand this difference in certainty around the age of four. They conducted an experiment in which children had to find a hidden candy. Two puppets gave them some information on the hidden candy. While one puppet told the child "I *know* it is in the red box", the other one said "I *think* it is in the blue box.". The child was then asked to look for the candy. They found that four-year-olds followed the more certain puppet's directions, indicating that the children understood the difference in speaker certainty expressed by the words *know* and *think*. The performance correlated with performance on ToM tasks. Theory of Mind was thus relevant for selective trust.

To find out whether this applied not only to selective trust in finding hidden objects, but also to selective word learning, Bergstra et al. (2013) investigated whether this difference in *know* and *think* could guide children in word learning, and whether it was again correlated to Theory of Mind. In their experiment, two puppets labeled novel objects. One puppet was certain, and said "I *know* this is a mit.". The other one was uncertain and said "I *think* this is a mit." It was found that 4- and 5-year-old children were significantly more likely to learn a new label from a certain puppet than from an uncertain puppet. Moreover, they tested three different ways of expressing speaker certainty. Besides the *know* vs. *think* (mental state verbs) condition, there was a condition in which the puppets expressed their certainty through describing familiarity with the object. One puppet stated that he played with the object all the time, while the other one said that he had never seen the object before. In a third condition, this description of familiarity was combined with the mental state verbs (the "both" condition). Although the children scored a little better in this last condition, this difference was not significant. It is thus clear from this study that speaker certainty is yet another factor that causes selective word learning. This study, however, failed to find a correlation with Theory of Mind performance of the children. This could be due to the small sample or to the scale of the Theory of Mind tasks which was only 4 points. However, there are other selective word learning studies that also fail to find a correlation with Theory of Mind (Lucas et al., 2013). Birch et al. (2008) try to explain this by suggesting that perhaps children do not really

consider the informants' mental states, but merely focus on their output. In accuracy studies such as theirs, the child would then just think "this informant said something bad", without considering what the informant knows, so without relying on mental state understanding. However, in a study such as Bergstra et al. (2013), none of the speakers provides "bad" output: *know* and *think* show a difference in certainty only. Still, it could be the case that the children do not really consider the speaker's mental states, and that it is only the hesitation of the speaker that makes them decide, rather than assumptions about the speaker's knowledge. However, Sabbagh & Baldwin (2001) conducted an experiment in which a knowledgeable speaker also showed signs of hesitancy. In the knowledgeable condition of this experiment, the speaker was knowledgeable about the label (a 'modi'), but hesitated which object he would want to call this, and said "I would really like to call one of these objects I made a modi, but I don't know which one". This reflected hesitancy, but not ignorance towards the label. In the ignorant condition, however, the speaker said "My friend says one of these toy she made is a modi, but I don't know which one", which shows ignorance. It was found that this hesitancy did not influence the performance of 4-year-olds (they still preferred to learn from the knowledgeable speaker), but it did influence the 3-year-olds. It thus seems that 4-year-olds do consider the speaker's mental state verbs, but 3-year-olds base their selective learning on the hesitancy of the speakers. The fact that a correlation between selective learning tasks and Theory of Mind tasks is present in some studies but absent in others, might thus be due to developmental differences between the children, small sample sizes, and other unfortunate consequences of experimental work. However, it might also be the case that even though mental states are considered by the children, this is not directly related to their ToM performance on false belief tasks. In the current study, Theory of Mind will be measured by means of the Theory of Mind scale (Wellman & Liu, 2004) which is a 5 point-scale and thus allows more room for individual differences. The possible relation between ToM and selective learning will be addressed again in the general discussion.

The current study will not only consider selective learning on the basis of speaker certainty, and find out if this is related to Theory of Mind, but will also investigate a more general reliability feature: a child's affect towards the speaker.

Kindness of the speaker

We have thus observed that children are not only sensitive to accuracy and knowledgeability of speakers, but also to general reliability features. Doebel & Koenig (2013) hypothesize that morality is one such feature, since morality might indicate that someone has good intentions and therefore has a general trustworthiness. They conducted an experiment in which children were either presented with a video including both a helpful informant and a neutral one (moral condition) or a video with a neutral informant and a harmful one (immoral condition). The harmful informant tore up drawings by his peers, while the helpful informant shared toys with his peers. The neutral informant never interacted with peers at all. After the video, the child was asked which of the two was nicer. Then the selective trust phase started. In this phase, there were four "ask trials", in which the child was asked which of the two informants she would like to ask for information. There were also four "endorse trials", in which the two informants named one object but each used a different label, and the child was asked what the object was called (the child's choice would indicate from which informant she had learned the new word). It was found that children were significantly better at distinguishing the harmful informant from the neutral informant than at distinguishing the helpful informant from the neutral informant. Apparently, negative behavior is easier to detect for children than positive behavior. Those children who correctly identified the informants were above chance in learning from the helpful informant rather than from the neutral informant, and in learning from the neutral informant rather than from the harmful informant. In short, the children who were able to point out which informant was nicer were also able to selectively learn from this nicer informant. This study thus showed that moral behavior of the speaker can cause selective learning in the learning of new words. Some issues are not yet entirely clear, however. For example what happens when moral and immoral behavior are contrasted directly. Moreover, what exactly "moral behavior" means in this context is not entirely clear. In the current study we therefore consider a similar, but more clearly defined feature: affect towards a speaker. We will investigate whether a child tends to learn from a speaker she likes (based on kind behavior of the puppet – similar to "moral behavior" of Doebel & Koenig) than from a speaker she does not like.

To sum up, recent research has shown that children are sensitive to many features of the interlocutor, such as accuracy, knowledgeability and morality, and that these features can cause selective word learning. The current study focuses on two interlocutor features: speaker certainty and affect. Three experiments were conducted, each one of which will be discussed

in detail below. The first experiment aimed to confirm the suggestion of earlier studies (e.g. Bergstra et al., 2013) that children are more likely to learn from certain speakers than from uncertain speakers. Experiment 2 aimed to investigate whether children are more likely to learn from speakers they like than from speakers they do not like, i.e. if affect can lead to selective word learning. Experiment 3 addressed these two factors (speaker certainty and affect) in interaction and aimed to find out whether speaker certainty or affect is guiding the child more, by contrasting kind but uncertain speakers with unkind but certain speakers. Furthermore, it contrasts kind and certain speakers with unkind and uncertain speakers to find out whether two interlocutor features combined will have a stronger effect (i.e., will lead to more selective word learning) than only one feature.

Two additional aims of the study are, first, to search for a relation between selective word learning and Theory of Mind (see above), which will be done by administering the Theory of Mind scale test (Wellman & Liu, 2004). Second, this study aims to address the concern of whether this kind of selective learning paradigm studies really focus on what could be "word learning" (Birch et al., 2008), or whether children do not just stick with one person who seems to provide better information. Therefore, it was also investigated whether children truly remember the labels of the objects that are used in the experiments.

Experiment 1

The main goal of experiment 1 was to find out whether children are more likely to learn new words from certain speakers than from uncertain speakers, which was investigated by means of a selective learning paradigm. Three additional tasks were administered: a Theory of Mind task (Theory of Mind scale, Wellman & Liu (2004)), to search for correlations between Theory of Minds skills and sensitivity to speaker certainty, a receptive vocabulary task (the Peabody Picture Vocabulary Test – III (Schlichting, 2005)), to search for possible correlations between current vocabulary and the ability to learn new words and to exclude children with very poor vocabulary (i.e. a standardized score below 70, 100 being the mean score), and "remember trials", in which it was investigated whether the child had really made new object – label connections.

Method

Participants

16 Dutch children (age range: 4;1 – 6;4, mean age 5;4) participated in the experiment. These were 4 4-year-olds (mean age 4;6), 9 5-year-olds (mean age 5;5) and 3 6-year olds (mean age 6;4). The children were recruited from three different classes of one elementary school.

Procedure

Each child was tested in a separate room in their school. All children participated in two sessions of approximately 20 minutes, with approximately 1 week in between the two sessions. The first session consisted of the main experiment and the Theory of Mind scale. The second session consisted of the Peabody Picture Vocabulary Test and a posttest investigating whether children had remembered the word labels. One adult (the experimenter) was present in both sessions. After the experiment, the children were rewarded with some stickers.

Main experiment

There were two different testing orders, each administered to half of the children, to prevent unwanted effects of order. The experiment started by introducing the child to two handpuppets; 'Green-y' and 'Blue-y' (two identical monkeys, one with a green scarf and one with a blue scarf). The experimenter told the child that the puppets would show objects and say what these objects were called. The experimenter also warned the child that she would not know all the objects, and that the puppets sometimes said things that were not right, so that the child must listen carefully. In the *practice phase* of the experiment, the puppets named toys that were familiar to the child. The practice phase was intended to let the children get used to the possibility that one puppet would be mistaken. Moreover, it was a check whether the child understood the procedure of the experiment and was capable of considering the utterances of the puppets before answering. One puppet said "This is a bike" and pointed at a bike, while the other puppet said "This is a bike" and pointed at a chair. The child was then asked "Which one is the bike? Point at the bike." Three similar items followed. Both puppets were correct half of the time. The order in which they spoke was also varied. After the practice phase, the experimenter said that the game was going to be a little more difficult, because the puppets would say words that the child would not have heard before. Then the *test phase* of the experiment started. Each child received 8 trials, in which the two puppets each named an unfamiliar object. One puppet was certain about this and said: "I have seen

this before. Look, this is how you pick it up," (and then picked up the object in front of him). "I have played with this a lot, because I have it at home, too. I know this is a *mit*." The other puppet was uncertain and said "I have never seen this before. I don't know how to pick it up" (and then picked up the object in front of him). "I have never played with this, because I don't have it at home. I think this is a *mit*." The puppets thus expressed their familiarity with the object and concluded with a mental state term (*know* or *think*) that described their level of certainty of the labeling event. The child was then asked the critical question: "Which one is the *mit*? Point to the *mit*." Choosing the object of the speaker who was certain would show that the child considered certainty of the puppets and on that basis displayed selective learning behavior. The experiment was balanced in the sense that each puppet was certain in 4 of the trials and uncertain in the 4 other trials. The labels of the items shown so far were repeated after 4 trials and at the end of the main experiment. 8 nonsense words (non-existent, but possible words in Dutch) were used in the experiment. The nonsense words were: *mit, klek, teg, glap, wop, prok, raf*, and *brim*. The toys used in the test phase were 16 home-made toys, which were all novel to the children.

Assessing Theory of Mind

The Theory of Mind scale tests were taken from Wellman & Liu (2004) and translated to Dutch. The Theory of Mind scale consists of five subtests on ToM which increase in difficulty. All five scenarios were administered consecutively, resulting in a maximum of five points if a child answered all questions correctly.

Assessing vocabulary

Receptive vocabulary of the children was measured by means of the Peabody Picture Vocabulary Test III (Schlichting 2005), a standardized picture selection task.

Remember-trials

To investigate whether children remembered the object labels, they were tested on 8 "remember trials". These remember trials were administered twice: once immediately after the main experiment and once in the second session, approximately 1 week after the first session. The remember trials were designed as follows: the experimenter showed the child the 16 objects again, two at the time as in the original trials, and asked for each set of two: "Which one was the X?" (using the nonsense word that was used in the main experiment). The child thus had to select which of the two objects corresponded to the newly learned label.

Results and discussion

Children's mean score in the main experiment, indicating the number of times that they chose the object of the certain puppet out of 8 trials, was 6.25 (M = 6.25, SE = 1.69, range 4 - 8). A one sample t-test showed that they scored significantly above chance level (t (15) = 5,316, p < .001). This indicates that children are more likely to learn a new label from a certain speaker rather than from an uncertain speaker, and confirms the idea that speaker certainty can cause selective word learning. No children were excluded from this study on the basis of their receptive vocabulary. No significant correlations with age, class or testing order were found, but there was a correlation with gender (r = .-595, p < 0.05). Girls (n = 9) scored better (M = 7.11, SE = 1.36, range 4 – 8) than boys (n = 7) (M = 5.14, SE = 1.46, range 4 – 8), and this difference was significant (t (14) = 2.774, p = .015). However, since (to my knowledge) no correlations with gender have been discussed in other studies in the field, it is unclear why this correlation was found. Possible correlations with gender will therefore also be investigated and discussed for experiments 2 and 3.

Scores on the ToM task ranged over the full scale, from 1 to 5 (M = 2.56, SE = 1.03), but did not show a significant correlation with sensitivity to speaker certainty. This point will be further discussed in the general discussion.

With regard to the remember trials, in the first remember trials, immediately after the experiment, children remembered correctly which object they had chosen before on average in 6.56 out of 8 trials (M = 6.56, SE = 1.15, range 4 – 8), which is significantly above chance level (t (15) = 8.891, p < .001). In the second remember trials, approximately one week after the experiment, the children chose the same object as before in 5.56 out of 8 trials (M = 5.56, SE = 1.41, range 3 – 8), which was also significantly above chance level, but significantly lower than one week earlier (t (15) = 3.303, p = .005). These scores show that the children seemed to remember some labels even after one week had passed (although less firmly than immediately after the experiment). This suggest that children really consider the label-object link, and do not just "stick with the accurate person" or follow some similar strategy.

Experiment 2

The main goal of experiment 2 was to find out whether children are more likely to learn new words from speakers they like than from speakers they do not like, i.e. to find out whether affect can lead to selective word learning. Similar to experiment 1, a receptive vocabulary task was administered Again, remember trials were added to the experiment, in which it was investigated whether the child had really made new object – label connections. The Theory of Mind scale was also administered, but is not included in the analysis, for no relation is expected between Theory of Mind and affect. The task was administered only to make sure the sessions were similar to those for the children of experiment 1.

Method

Participants

15 Dutch children (age range: 4;0 – 6;5, mean age 5;3) participated in this experiment. These were 5 4-year-olds (mean age 4;8), 8 5-year-olds (mean age 5;4) and 2 6-year olds (mean age 6;4). The children were recruited from three different classes of one elementary school.

Procedure

The procedure was similar to that of Experiment 1; all children participated in two sessions of approximately 20 minutes, with approximately 1 week time between the sessions.

Main experiment

The puppets' introduction and *practice phase* were identical to that of Experiment 1. After the practice phase, the *"affect-phase"* followed. The experimenter said that she was going to give some stickers to Green-y and Blue-y. Both puppets received some stickers and reacted to this. One puppet recognized that the child did not receive any stickers, and shared his stickers with the child, thereby displaying kind behavior. The other puppet was not willing to share the stickers and said that he would keep the stickers all to himself, thereby displaying unkind behavior. For half of the children, Blue-y was the kind, sticker-sharing puppet, for the other half Green-y was the kind, sticker-sharing puppet. Next, the experimenter told the child that the game was going to be a little more difficult, because the puppets would say words that the child would not have heard before. Then the *test phase* of the experiment started. Each child received 8 trials, in which the two puppets each named an unfamiliar object. They both had an object in front of them and said, for example "This is a *mit*". Since what they said was identical, the only reason for the child to distinguish between the two would be to rely on the

previous *affect-phase*. After the labeling, the child was asked the critical question: "Which one is the *mit*? Point to the *mit*." Choosing the object of the speaker who had shared the stickers would show that the child showed selective learning on the basis of affect. The order in which the puppets spoke was varied. After the 8 trials, the child was asked which puppet she found nicer, Green-y or Blue-y, to make sure the child distinguished between the two. The labels were identical to those of Experiment 1 and were repeated after 4 trials and at the end the main experiment.

Assessing vocabulary

As discussed for experiment 1, receptive vocabulary of the children was measured by means of the Peabody Picture Vocabulary Test III. One child from originally 16 participants was excluded from data analysis because of scoring 48, which translated to a score below 70 for the standardized score, leaving 15 participants for analysis.

Remember-trials

Similarly to Experiment 1, remember trials were administered twice.

Results and discussion

Children's mean score in the main experiment, indicating the number of times that they followed the nice, sticker-sharing puppet out of 8 trials, was 4.47 (M = 4.47, SE = 1.73, range 1 - 8). A one sample t-test showed that this was not significantly above chance level (t (14) = 1,047, p = .313). 3 children did not chose the sticker-sharing puppet as the nicer puppet, but excluding those children from analysis would actually result in a lower mean score of 4.42. The fact that children do not follow the nice puppet significantly more than chance suggests that affect cannot cause selective word learning. However, the result is not extremely convincing, since this was a small sample and some of the children actually did follow the nice puppet more than half (chance-level) of the time, especially the younger ones. It could thus be that affect only plays a role for younger children, or that it does not play a big role on its own, but it could have a small, additive effect on speaker certainty, as will be investigated in Experiment 3.

No significant correlations with age, gender, class or testing order were found, nor with receptive vocabulary scores (M = 75.47, SE = 15.06, range 48 – 102). Again, this could be

due to the small sample size and it is expected that correlations might be found with age in Experiment 3, where a larger sample is analyzed.

With regard to the remember trials, in the first trials, immediately after the experiment, the mean score (indicating the times the child chose the same object as she had chosen before) was 7.00 out of 8 trials (M = 7.00, SE = 1.00, range 5 − 8), which is significantly above chance level (t (14) = 11.619, p < .001). In the second remember trials, approximately one week after the experiment, the mean score was 5.60 out of 8 trials (M = 5.60, SE = 1.35, range 4 − 8), which was also significantly above chance level (t (14) = 4.583, p <.001), but significantly lower than one week earlier (t (14) = 3.309, p = .005). This result is similar to that of Experiment 1: children seemed to remember many labels even after one week had passed (although less than immediately after the experiment). This again suggests that children really consider the label-object link, and do not just "stick with the accurate person" or adopt some similar strategy.

Experiment 3

Experiments 1 and 2 investigated the effect of speaker certainty and affect on word learning. The main goal of experiment 3 was to find out what happens when speaker certainty and affect interact. It was designed to answer two questions. The first is: when both affect and speaker certainty come into play, which factor guides the child more in word learning? Based on the results of Experiment 1 and 2, it was expected that speaker certainty rather than affect would guide the children. However, if this is the case, affect could still have an additive effect: the second question is therefore whether these factors can reinforce one another (i.e., if a speaker is certain and kind, will this guide the children more than when he is either kind or certain only?). Similar to Experiment 1 and 2, the Theory of Mind and Vocabulary test as well as the remember trials were administered.

Method

Participants

Originally, 52 children participated in the experiment, but 3 had to be excluded because of missing data. This left 49 Dutch children (age range: 4;0 − 6;7, mean age 5;4) in the experiment. These were 20 4-year-olds (mean age 4;8), 16 5-year-olds (mean age 5;7) and 13 6-year olds (mean age 6;3). The children were recruited from four different classes of two elementary schools.

Procedure

The procedure was similar to those of Experiments 1 and 2; all children participated in two sessions of approximately 20 minutes, with approximately one week time in between the sessions.

Main experiment

The experiment was a combination of Experiments 1 and 2: both affect (in the *affect phase*) and speaker certainty (in the *test phase*) were introduced as interlocutor features. The introduction of the puppets and the *practice phase* were identical to those of Experiments 1 and 2. After the practice phase, the *"affect-phase"* followed, in which one of the puppets shared stickers, similarly to Experiment 2. In the *test phase* each child was presented with 12 trials, in which the two puppets each named an unfamiliar object. Similar to Experiment 1, one puppet was certain, one puppet was uncertain. The puppets expressed their familiarity with the object and concluded with a mental state term (*know* or *think*) that described their level of certainty of the labeling event. The child was then asked the critical question: "Which one is the *mit*? Point to the *mit*.". This combination of affect (introduced in the *affect phase*) and speaker certainty (at each trial in the *test phase*) led to two different kinds of trials. In half (six) of the trials, a kind and uncertain puppet was contrasted with an unkind and certain puppet ("contrast 1"). These trials could thus show whether affect or speaker certainty would guide the children. In the other six trials, a kind and certain puppet was contrasted with an unkind and uncertain puppet ("contrast 2"). These trials could thus show whether affect and speaker certainty could reinforce each other. Similar to Experiment 2, the experiment asked which puppet the child found nicer. The 12 nonsense words used in this experiment were: *mit, klek, hast, teg, glap, virg, wop, prok, tork, raf, brim* and *nelf*.

Assessing vocabulary

Receptive vocabulary of the children was again measured by means of the Peabody Picture Vocabulary Test III (Schlichting 2005).

Remember-trials

12 remember trials (similar to Experiments 1 and 2) were administered twice.

Results and discussion

To analyze the results, the trials were split into two categories: the "contrast 1 trials" (kind and uncertain puppets vs. unkind and certain puppets, which aimed to find out whether affect or speaker certainty would have a greater guiding effect in word learning), and "contrast 2 trials" (kind and certain puppets vs. unkind and uncertain puppets, which were designed to show whether affect and speaker certainty could reinforce each other). Regarding the "contrast 1 trials", children on average chose the object of the unkind and certain puppet 4.18 times out of 6 trials (M = 4.18, SE = 1.62, range 0 - 6). The children followed the unkind and certain puppet significantly above chance level (t (48) = 5.13, p < .001). This shows that when both speaker certainty and affect come into play, it is speaker certainty, rather than affect, that causes selective word learning by the children. Despite this, it could be the case that affect still has an effect on word learning. It could, for example, add to the effect that speaker certainty already has: kind and certain puppets could be followed even more than puppets who were certain but not kind. The "contrast 2 trials" were trials in which a kind and certain puppet was contrasted with an unkind and uncertain puppet. Children on average followed the kind and certain puppet more often than the unkind and uncertain puppet (M = 4.51, SE = 1.49, range 1 – 6) and did this significantly above chance level (t (48) = 7.106, p < .001). This was expected, but it would be interesting to see if this effect is bigger than when selective learning occurs on the basis of speaker certainty *only*. A paired samples t-test shows that children follow the certain puppet in contrast 2 trials (i.e. the puppet that was also kind) significantly more than the certain puppet in contrast 1 trials (i.e. the puppet that was also unkind) (t (48) = 2.367, p = .022). This difference shows that children are not indifferent to affect; it adds to the effect of speaker certainty; certain speakers that are also kind cause a bigger effect of selective learning than speakers that are not kind.

Surprisingly, only 27 out of 49 children regarded the kind, sticker-sharing puppet as the kinder puppet. The other 22 children chose the unkind puppet, or both puppets, as being the kinder one. On the basis of Experiment 2, in which 13 out of 16 children recognized the sticker-sharing puppet as the kind puppet, this was not expected. Raw data of only the 27 children that preferred the sticker-sharing puppet show that the means are a little bit higher for this group (M = 4.41 for contrast 1 and M = 4.81 for contrast 2). T-test were performed on this sample of children, which showed that the results were similar to those of the whole group. In the "contrast 1 trials", the children chose the object of the certain and unkind puppet (rather than that of the uncertain and kind puppet) significantly more than chance (t (26) =

5.146, p < .001). In "contrast 2" trials, they chose the object of the certain and kind puppet (rather than that of the uncertain and unkind puppet) significantly more than chance (t (26) = 6.663, p < .001) These results and the fact that no correlation is found between the children's scores and their puppet's preference show that the children who did not prefer the sticker-sharing puppet did not perform very differently from the other children. Perhaps, the children did not *consciously* prefer the sticker-sharing puppet, but were still influenced by the sharing-action of this puppet, and therefore more likely to learn a new word from this puppet. Moreover, the "which puppet do you find nicer?"-question was asked after the experiment; children could have changed their opinions by then. The puppets were also present at this question, which may have led children to choose both puppets, to avoid "hurting the other puppet's feelings". In any case, the fact that only 27 children recognized the sticker-sharing puppet as being kind shows that this might not have been the best way to introduce a kind and an unkind speaker. This point will be addressed in the general discussion.

Since the results of Experiment 2 suggested that 4-year-olds might be more sensitive to an effect of affect than older children, data of 4-year-olds (n = 20) were re-analyzed separately. A t-test showed that 4-year-olds on average follow the certain and kind puppet in the "contrast 2 trials" in 4.15 out of 6 trials (M = 4.15, SE = 1.50, range 2 − 6), which was significantly more than chance (t (19) = 3.437, p = .003). However, in "the contrast 1 trials", they score at chance level (M = 3.60, SE = 1.50, range 0 − 6). Similar to the results of the full sample, there is a difference between 4-year-olds scoring on "contrast 1 trials" and "contrast 2 trials" (they chose the object of the certain speaker more in "contrast 2", when the certain speaker is also kind) (t (19) = 2.463, p = .024), but the effect of affect does not seem to be bigger for 4-year-olds.

No significant correlations with class or testing order were found, but all tests correlated with age, indicating the developmental nature of the tests. A significant correlation between the number of times the object of the certain speaker was chosen and Theory of Mind (r = .556, p < .001 for contrast 1 and r = .487, p < .001 for contrast 2) was found this time (as opposed to Bergstra et al., 2013, and Experiment 1), suggesting that there does exist a link between ToM skills and sensitivity to speaker certainty (but see the General Discussion for further discussion). A moderate correlation that is harder to explain is the correlation between gender and amount of times the object of the certain speaker was chosen (r = .301, p = .036 for contrast 1, r = .285, p = .047 for contrast 2). The correlation, however, was in the opposite

direction from that in Experiment 1: this time boys scored higher at following the certain puppet than girls (boys: M = 4.70 for contrast 1 and M = 4.96 for contrast 2, girls: M = 3.73 and M = 4.12). These contrasting findings makes it hard to draw strong conclusions about possible gender differences. Moreover, the difference between boys and girls in this experiment was probably due the fact that boys in this sample were on average older (mean age 5;8) than the girls (mean age 5;2).

Regarding the remember trials, in the first trials, immediately after the experiment, the mean score (indicating the times the child chose the same object as she chose before) was 9.84 out of 12 trials (M = 9.84, SE = 1.77, range 6 − 12), which is significantly above chance level (t (48) = 15.158, p < .001). In the second remember trials, approximately one week after the experiment, the mean score (indicating the times the child chose the same object as she chose before) was 8.90 out of 12 trials (M = 8.90, SE = 1.87, range 4 − 12), which was also significantly above chance level (t (48) = 10.827, p <.001), but significantly lower than one week earlier (t (48) = 3.951, p < .001). There was a correlation between the first remember trials and the times the object of the certain speaker was chosen in the word learning part of the experiment (r = .651, p < .001 for contrast 1 trials, r = .546, p < .001 for contrast 2 trials). This correlation was also present for the remember trials after one week (r = .378, p = .007 for contrast 1 trials, r = .460, p = .001 for contrast 2 trials), suggesting that the children who showed more selective learning on the basis of speaker certainty were better at remembering the words.

General discussion

This study has considered two factors that might influence a child's word learning: speaker certainty and a child's affect towards the speaker. It was found that speaker certainty guided children in word learning, but affect on its own did not. Speaker certainty was indeed shown to overrule affect in word learning; children were more likely to learn a new word from a certain and unkind puppet than from an uncertain and kind puppet. However, affect does have an additive effect on selective learning: comparing these trials to those in which a kind and certain puppet was contrasted with an unkind and uncertain puppet showed that in the latter type of trials, children followed the certain puppet even more. In other words; children followed the certain and kind puppet more often than they followed the certain and unkind puppet. Affect can thus also influence selective word learning, even though its guiding effect is smaller than that of speaker certainty.

One aspect that could have influenced the results was the fact that, in Experiment 3, only 27 out of 49 children regarded the sticker-sharing puppet as the kinder puppet. Although exclusion of the other children in the analysis did not alter the findings, it is possible that affect plays such a small role, exactly because the children did not feel very much affect towards any of the puppets during the experiment. Perhaps sticker-sharing is not the best action to present speakers as kind and unkind. One could even imagine that children themselves in daily life are not willing to share stickers with others, so it might not be a reason to really dislike someone (and, as a consequence, to regard him as unreliable in his word labeling). Future research could try to use other settings in which the kind – unkind contrast is more clear, and investigate if the effect of affect would be bigger in that case.

This study has also aimed to address the worries towards studies in this field (Birch et al., 2008), asking whether these studies can really be called "word learning", and whether children do not just stick with one person who seems to provide better information. After all three experiments, children were presented with remember-trials, which investigated whether the children recalled which of the two objects matched with the newly learned label. It was found that children did remember many of the object-label links. The link between object and label was still present after one week, even though the speakers (puppets) were not there at that time, which suggests that children really learn a new object–label link, and do not only focus on which speakers seems to provide better information. Based on these results and those of Birch et al. (2008), we claim that studies using these paradigms can indeed be called "word learning studies".

Interestingly, this study found a correlation between Theory of Mind and sensitivity to speaker certainty (i.e. performance in the main part of Experiment 3), but it was found only in Experiment 3, not in Experiment 1. This could be due to a difference in sample sizes, but the fact that correlations between Theory of Mind and selective learning continue to show up in some studies but not in others (Lucas et al., 2013) suggests that there is more to this unclear (lack of a) relation. As Bergstra et al. (2013) suggested that even though false belief tasks (and other Theory of Mind tasks) and understanding of speaker certainty are both related to beliefs, they may tap into different aspects of children's cognitive developments. The fact that the correlation sometimes does show up may just be due to coincidence. Children may sometimes be developing Theory of Mind capacities and understanding of speaker certainty at the same time, without there being a relation between the two.

To conclude, this study has contributed to the field of selective (word) learning by showing that speaker certainty is an important cue in word learning and that affect can add to this, but does not guide children on its own. This study adds to recent findings showing that children consider several kinds of cues from the interlocutor in word learning, ranging from prior accuracy to more general reliability features, and highlights the fact that word learning is not only a cognitive, but also a social process.

References

Bergstra, M., De Mulder, H.N.M. & Coopmans, P. H. A. (2013). Children's ability to use speaker certainty in learning novel words. In: Aalberse, S. & Auer, A. (Eds.) *Linguistics in the Netherlands 2013* [AVT 30] (pp. 1-12). Amsterdam, John Benjamins Publishing Company.

Birch, S. A.J., Vauthier, S.A. & Bloom, P. (2008). Three- and four year olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018– 1034.

Corriveau, K.H., Kinzler, K.D. & Harris, P.L. (2013). Accuracy trumps accent in children's endorsement of object labels. *Developmental Psychology, 49*, 470 – 479.

Jaswal, V. K., & Neely, L.A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17*, 757–758.

Koenig, M. A., Clément, F., & Harris, P.L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Koenig, M. A. & Sabbagh, M.A. (2013). Selective social learning: New perspectives on learning from others. *Developmental Psychology, 49*, 399 – 403.

Lucas, A. J., Lewis, C., Pala, F.C., Wong, K. & Berridge, D. (2013). Social cognitive processes in preschooler's selective trust: Three cultures compared. *Developmental Psychology, 49*, 579 – 590.

Markman, E. M. (1994). Constraints children place on word meanings. In P. Bloom, ed., *Language acquisition. Core readings.* Cambridge, MA.: MIT Press, 154-173.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology, 49,* 404–418.

Moore, C., Angelopoulos, A. & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60 – 68.

Moore, C., Pure, K. & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development, 61,* 722–730.

Sabbagh, M. A. & Baldwin, D.A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL.* Amsterdam: Harcourt Test Publishers.

Sobel, D.M. & Macris, D.M. (2013). Children's understanding of speaker reliability between lexical and syntactic knowledge. *Developmental Psychology, 49*, 523 – 532.

Sobel, D.M & Corriveau, K.H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*, 669 – 679.

Wellman, H.M. & Liu, D. (2004). Scaling of Theory-of-Mind tasks. *Child Development, 75*, 523–541.