Learning verbs more effectively through meaning congruent action animations

Lea A. Hald a, *, Marianne van den Hurk b, Harold Bekkering a

a Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, PO. Box 9010, 6500 GL Nijmegen, The Netherlands
b Pedagogical Sciences, Education Science, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

ABSTRACT

The current study investigates the effectiveness of learning words while displaying meaning congruent animations. We explore whether learning words with animation is sensitive to properties known to influence action understanding. We apply an embodied cognition framework and predictions from a recent theory about language and action (Action-Based Language theory, Glenberg & Gallese, 2012). The current study aims to investigate whether dynamic animations add to word learning (Experiment 1) and what the linguistic relation between the dynamic animation and the word learning is (Experiment 2). Results indicate that meaning congruent animations improved verb learning compared to meaning incongruent animations when measured by a recognition task. When measured by an active recall task, congruent animations led to better learning than static pictures. In both measures, meaning congruent animations support word learning. Experiment 2 replicates and extends this and suggests that highlighting conceptual information related to the dynamic action (such as the goal) improves word learning further. The findings are in line with Action-Based Language theory, which suggests that children are able to make better simulations of an action during learning when supported by meaning congruent animations. Highlighting conceptual information additionally supports this learning process.

© 2015 Elsevier Ltd. All rights reserved.

1. Animations and word learning

Much of the research using dynamic multimedia for word learning involves gesture, which has to be imitated by the participant, or the participant has to do some other type of movement. An exception to this is seminal work by Plass, Chun, Mayer, and Leutner (1998, 2003). They used multimedia (video recordings and pictures) to explore aspects of individual differences related to learning and cognitive load. Using an interactive multimedia program, American college students learning German as a second language were presented with a short story in German. For a selection of key words in the story, students could access an annotation about the meaning of that word. The annotations were either verbal or visual, where the visual condition contained either pictures or short video recordings. In Plass et al. (1998) the students choose which type of instruction they received while in Plass et al. (2003) the annotation mode was predetermined between subjects. Overall, there were improvements in recall of individual vocabulary items when students used both visual and verbal annotations (Plass et al., 1998) but this was dependent on individual differences in spatial and verbal abilities (Plass et al., 2003). However, most
important for the current study, Plass et al. did not report on the difference in learning between still pictures and short video recordings. Video recordings and pictures were grouped together in the visual annotations.

Results from other research using videos or animations for vocabulary learning have been mixed. For example, Neuman and Koskinen (1992) explored the effectiveness of watching television with or without subtitles in comparison to reading and listening to the same information or simply listening. They found the greatest gains in learning when participants watched subtitled television, suggesting that animations can aid word learning when presented together with the word. More recent research looking at vocabulary learning from watching educational television showed similar gains in learning words (Linebarger, Moses, Liebeskind & McMenamin, 2013). However, the effectiveness of subtitles (displaying the word to learn) varied based on the child’s socioeconomic status: Low socioeconomic status children were not helped by the subtitles. Middle-class socioeconomic status children were helped by the subtitles, but the subtitles did not lead to additional gains when there was repeated exposure (Linebarger et al., 2013).

Research by Baltova (1994) used either combined video-audio clips or audio-only clips to teach French to English speaking 8th-grade students and found no significant difference. Furthermore, Sun and Dong (2004) found animation in itself was ineffective for teaching English to Chinese children. Only when the animations were paired with textual support did they find improvements in learning. Other studies using video and subtitles have shown that subtitles in a second language can aid perceptual learning (Mitterer & McQueen, 2009).

In summary, meaning congruent video or animations may help with word learning, but it is unclear whether animations can provide additional benefits compared to pictures. It should be noted that none of the research above was intended to explore the added benefit that meaning congruent animations may bring over meaningful congruent pictures, but this research focused mainly on visual versus verbal information instead (e.g., Plass et al., 1998).

There is a long tradition of exploring the use of animations versus pictures for learning other types of information, such as procedural motor knowledge, declarative knowledge or problem solving (see Höffler & Leutner, 2007 for a review). In their meta-analysis Höffler & Leutner (2007) conclude that representational animations improve learning more than decorative animations. Furthermore they suggest that learning procedural motor knowledge benefits the greatest from animation, followed by learning declarative knowledge and then problem solving. The largest improvement with learning procedural motor knowledge with animations could be due to the fact that these animations illustrated motor actions. This would be in line with our hypothesis about word learning via motor actions. There is some recent support for this possibility. For example, Brucker, Ehlis, Häußinger, Fallgatter and Gerjets (2015) found that university students could learn patterns of fish movements’ best when they watched animations of hands gesturing the correct movements. Similarly, it has been suggested that learning biological or manipulative tasks may be better learned by animations, where learning non-biological or symbolic tasks may be better learned by static pictures (Castro-Alonso, Ayres & Paas, 2014). This suggestion makes it particularly interesting to investigate where learning word meanings fit in this continuum.

None of the research on using animations for word learning systematically used animations that demonstrated motor actions. However, there is an independent reason to believe that watching motor actions may be useful for word learning. The embodied cognition framework and in particular, a new theory specifically outlining the coupling of language and action (Glenberg & Gallesse, 2012), suggests that perceiving congruent motor actions is important for word learning.

1.1. Motor actions and embodied cognition

Why might motor involvement in general help with word learning? One explanation is given by the theory of embodied cognition: The basic idea is that our knowledge about objects and other referents in the world, and the way in which we understand the world, is grounded in our perception and actions, rather than some type of abstract symbols (Barsalou, 1999). For example, our concept of bird is based on the collection of bodily states (perceptions, actions and emotions) we have experienced with birds. Broadly speaking, these previous perceptions are partially reactivated (simulated) in order to understand bird. This reactivation or simulation is not considered to be a conscious mental image; instead simulation is more like a record of previous neural states (see Barsalou, 1999; Taylor & Zwaan, 2009; Glenberg & Kaschak, 2002). While it provides us with an interesting starting point, the framework of embodied cognition does not necessarily lead to any specific predictions about the use of meaning congruent animations compared to pictures for word learning.

1.2. Action-Based Language theory

Recently a formal theory has been proposed to account for the relationship between the action aspect of embodied cognition and language, the Action-Based Language theory (ABL, Glenberg & Gallesse, 2012). The ABL theory includes a mechanism of motor control and paired controller/predictor models to account for language learning, comprehension and production. The ABL theory makes use of two well-established models of motor control as the framework for the relationship between action and language, the MOSAIC and HMOSAIC theories (Haruno, Wolpert, & Kawato, 2003). These models propose a controller and a predictor. The controller computes motor commands from a representation of goals and context, while a predictor predicts the motor and sensory consequences of actions and sends error information to the controllers to update motor commands as needed.

What is novel about the ABL theory is that Glenberg and Gallesse propose that the brain solves the problem of how to control contextually-appropriate action and contextually-appropriate language via the same process. They propose that the opportunistic sharing of action control and language processing is possible through mirror neurons. The theory links verb meaning directly to the motor actions associated with the word via the controllers. In addition, meaning is linked directly to the predictions (from the predictors) that can be derived from simulating the action. The ABL theory predicts that meaning congruent animations would allow the child to encode the word through a simulation so that better predictions can be made. For example, if learning the word chiseling, a simulation would lead to predictions about how a chisel can be used, what would result from using a chisel, and what type of objects would afford chiseling. This type of information could be predicted much better when viewing an animation. Therefore, encoding the word assisted by animation should lead to a more comprehensive understanding of the word meaning.

In sum, according to the ABL theory, understanding what it means to chisel involves the sensory-motor feedback of chiseling. This predicts that performing the action yourself would provide the strongest environment for this learning to occur. But the difference...
here would be one of degree: Sentences about an action, pictures of an action and animations of the required action should all provide some amount of context and allow for better predictions to be made while encoding the word.

This all assumes that the animation demonstrates the use of the new word, which we call meaning congruent animations here. If the animation shows a person using the chisel for an unrelated action (like putting it away) this is a meaning incongruent animation, which should reduce or even reverse the positive effects.

1.3. Action-based word learning

Evidence for the ABL theory of word learning comes mostly from research using imitation, enactment or gestures. It has been demonstrated that imitating meaning congruent gestures shown in videos can aid second language learning in adults (compared to imitation, enactment or gestures. It has been demonstrated that imitating meaning congruent gestures shown in videos can aid second language learning in adults (compared to demonstrating that imitating meaning incongruent gestures; see Macedonia, Müller, & Friederici, 2011; Schmidt-Kassow, Kulka, Gunter, Rothermich, & Kotz, 2010; Macedonia, 2003). Similarly, children have been shown to remember more items from a list with concrete nouns and descriptive adjectives in a second language when they imitated meaning congruent gestures that accompanied the words during training, than when they only observed these gestures (Tellier, 2008; see also Allen, 1995).

Related to this are findings about improved memory for words when meaning congruent actions are self-generated during training, a phenomenon that is called the enactment effect (see review by Engelkamp, 1998; Nilsson, 2000). For example, memory for both concrete and abstract words is improved when adults are cued during recollection with gestures that they made for the to-be-remembered words, compared to when they were not cued with gestures, or cued with someone else's gesture (Frick-Horbury, 2002).

Performing meaning congruent actions is not necessarily better than only observing those actions. Feyereisen (2009) asked participants to read phrases containing action verbs (e.g., to peel a potato). Some participants were additionally asked to watch a congruent gesture (e.g., observing someone peeling a potato) or imitate the action (e.g., pretend to peel a potato themselves). A facilitative effect on memory was found for both action observation and action imitation, compared to word reading alone. However, there was no difference between the action observation and action imitation (Feyereisen, 2009).

1.4. The current study

The current study investigates word learning in children 7–8 years old. Previously it has been demonstrated that word learning can be improved by showing a picture along with text (Smith et al., 1987; Smith et al., 1994). Related to this, watching a video of a meaning congruent gesture along with imitating the gesture also improves word learning (Macedonia et al., 2011; Schmidt-Kassow et al., 2010; Tellier, 2008; Allen, 1995). Here, we directly compare word learning by simply watching a meaning congruent dynamic action versus a meaning congruent picture (without performing an imitation/action) when it is paired with a meaningful context sentence. This question is of theoretical interest and of direct relevance to future teaching practice: Textbooks can easily reproduce informative pictures but a multimedia setup is required for animations. However, if dynamic animations add to novel word learning, creating the learning infrastructure to employ the use of animations is desirable.

According to the ABL theory, verb meaning is directly linked to motor actions associated with a word and the predictions that can be derived from simulating an action. Meaning congruent animations, by nature of the dynamic information provided about the action should allow for the child to encode the word and accompanying sentence so that better predictions can be made, which should lead to a better understanding of the words’ meaning (Glenberg & Gallee, 2012). It is the interdependent effect of the meaning congruent animation along with the accompanying sentence that should lead to better learning.

To test this hypothesis, Experiment 1 includes three different training conditions for learning the meaning of 21 existing verbs that are unfamiliar to the children tested. The between-participant conditions were as follows: 1) each verb paired with a meaning congruent animation (MCA); 2) each verb paired with a meaning congruent picture (MCP); 3) each verb paired with a meaning incongruent animation (MIA). MIA was a control condition for the effect of an animation in itself (see Methods section for details of the control condition). Regardless of the visual context, the same sentence using the verb correctly was paired with the verb.

Experiment 2 was conducted in part to replicate the findings of Experiment 1. Moreover, Experiment 2 is designed to strengthen the argument that it is the pairing of the sentence to a meaning congruent animation that leads to improved simulation, which in turn enhances the understanding of the word.

2. Methods

2.1. Participants

A total of 168 children in 2nd grade from a number of primary schools across several cities in the Netherlands participated in the experiment at their school. Of these, 156 are included in the final analysis (72 females; aged 7-8, mean = 7.3; 122 native Dutch speakers, 34 non-native Dutch speakers). As we discuss in the Results section, there were no significant differences between the performance of native and non-native Dutch speakers on the posttests. The 156 participants were allocated randomly across the three conditions resulting in 51 participants in MCA, 54 in MIA and 51 in MCP. Twelve participants (of the original 168) were excluded from the final analysis because of illness at one session (6 children) or experimental error (6 children). The experimental error involved accidentally skipping a page from the essential word pretest (see Pretest below). This meant that we have no information whether words were indeed unfamiliar for 38% of the data, so we did not analyze these children's data further.

2.1.1. Sampling

Twenty-five schools close to our university were sent a letter describing the research project. A week later we called the schools to ask if they were interested in participating. Schools that participated received €50 to buy books for their library. Only children in 2nd grade were asked to participate in the study.

2.1.2. Stimulus materials

Twenty-one verbs were chosen for children to learn. All verbs denoted an action that was concrete and could be visualized (e.g., to chisel, to sketch) and therefore easily animated. The verbs selected had either appeared in a language subsection of a Dutch national standardized test1 between 2004 and 2006 or came from the Schrooten and Vermeer (1994) list of words that children encounter in primary school. Linguistic factors, such as word length, number of syllables and word frequency were kept as similar as possible across words. Using SUBTLEX-NL, a database of

---

1. This test is administered to 12 years olds (Cito Eindtoets Basisonderwijs) in school.
frequency of use of words in Dutch (from Keuleers, Brysbaert, & New, 2010), the average log frequency was .54 (maximum = .96; minimum = 0). The average number of syllables was 2.86 (maximum = 4; minimum = 2). The average length in letters was 8.62 (maximum = 12; minimum = 6).

The training materials for each word for the three conditions resulted in two video animations for each verb and one picture for each verb. Each animation lasted 4 s. For the meaning congruent animations (MCA) an avatar illustrates the necessary specific movement(s) needed to perform the action (verb) on a particular object correctly. The meaning congruent pictures (MCP) were created by making a still shot of the MCA at the point in time the avatar acts upon the relevant object. The meaning incongruent animations (MIA) were created by switching the movement information and related posture of the avatar across verbs. For example, in order to create the MIA for the verb to chisel the movement that we used for the verb to hoe was exchanged with the movement used in to chisel (see Fig. 1 for example). This is intended to be a control for the effect of movement per se. What this means in practice is that the MIA for chiseling involved the avatar holding the chisel in a standing position above the stone and moving the chisel in a hoeing movement. The MIA for hoeing involved the avatar holding the hoe above grass but using a chiseling movement, which looks like moving the hoe up and down on top of one point in the grass. The MCA and MIA are perceptually similar since both contain dynamic movement with the same correct objects. The object held by the avatar (e.g. chisel) implies certain functional entailments, which may in itself prime the actions. However since the objects are held constant across the three conditions (even in the MIA condition) this cannot contribute to the difference in conditions we expect to find.

Although perceptually very similar, the key difference between the MCA and MIA conditions is the goal-directedness of the action. In the MCA, there is a better fit between the affordance of the object and how effective the action is in changing the environment. This is apparent even when the action does not actually fully change the environment (for example in bouncing a ball, neither the ball or ground change appearance after correctly bouncing a ball, but making the correct hand movements for bouncing is an effective action to make with a ball).

There is already an extensive literature about the differences between learning words with text only versus text plus pictures. Feyereisen (2009) specifically showed that imitation of actions leads to better memory than text alone, but they did not include a picture only condition. We therefore felt that adding a text only condition would not be very informative as a comparison condition, as it would certainly perform less well than the two conditions of interest (MCA, MCP).

There are still other possibilities for the control condition, such as a series of still pictures, or a still picture taken from the MIA. However, the design of this study was modeled after word learning studies using gesture (e.g., Macedonia, et al. 2011), which are most alike to this field of research. The designs of these studies typically involve a congruent movement gesture video and an incongruent or meaningless movement gesture video to control for the effect of movement in itself. The MIA condition was made to be similar to the incongruent/meaningless gesture videos that are typically used.

Related to this point, the choice to call these animation conditions meaning congruent and meaning incongruent (MCA and MIA) does not imply that we believe we are actually manipulating congruence. Since there is only cultural convention that connects these verbs with their referents, there is no reason in principle that ‘chiseling’ is not something that you do with a long tool in water for example. Therefore we do not believe that we are directly manipulating congruency per se. One way to think about the incongruent animations is that the action that is performed does not appear to change anything in the animations, whereby the congruent animations at least suggests that the action illustrated could lead to a change in the object present. The use of the terms meaning congruent and incongruent were chosen based on terminology to describe similar manipulations in the gesture literature (a meaningful versus meaningless gesture; e.g., Macedonia, et al. 2011). Notably participants in Macedonia et al. (2011) are given a translation of the word they are supposed to know paired with a

---

**Fig. 1.** Example training materials Experiment 1. Images from the videos and picture used for the three training conditions: (A) meaning congruent animation, (B) meaning congruent picture and (C) meaning incongruent animation. The videos all contained the same avatar performing the movements. The verb to learn was played aloud followed by a sentence using the verb. The example here shows to chisel (beitelen).
congruent or incongruent action, where we have not given a translation or definition. At the same time, the sentences that accompany the animations do use the verb correctly, and they do give a clue as to the meaning of the verb, as opposed to presenting the verbs paired with animations in isolation. Furthermore, even though there is no inherent conflict between what is said in the context sentence and what is shown in the incongruent animation, the sentences make more sense (are more congruent) with the congruent animation. Since our experiment is largely based on the gesture literature that uses this terminology, it made sense to us to use comparable terminology, even though this choice is not fully satisfactory.

To ensure that our animations depicted information that is informative about the verb and to see how engaging the animations were, we conducted a ratings test. Thirty-two adults (21 females; mean age 25.2) rated the animations for course credit or a small monetary reward. Each participant watched the animations or pictures (without hearing the accompanying sentence) from only one of the three conditions. They were asked to describe what they saw in as few words as possible and to rate on a 5-point scale (5 was most interesting) how interesting they thought the animation was. For the MCA, participants used the exact word we intended to illustrate 29% (e.g., beitenlen - chiseling) or a related word(s) 35% (e.g., steen bewerken — working with stone) of the time. For the MIA, participants used the exact word 7% of the time and a related word(s) 22% of the time. Finally for the MCP the participants used the exact word intended 16% of the time and 33% of the time a related word(s). There was a significant difference between the ratings given across the three conditions (F (2, 669) = 15.3, MSe = .76, p < .001; \( \eta^2_g = .04 \)). The ratings for MCA matched the exact word or related word more often than the responses for MIA (t (439) = 5.3, p < .001, d = .51). Similarly the rating for MCP matched the exact word or related word more often than the responses for MIA (t (439) = 4.3, p < .001, d = .41). There was not a significant difference between the ratings for MCA and MCP (p = .40, d = .08).

In terms of engagement or interestinosness of the animations there was a significant difference between the ratings given across the three conditions (F (2, 669) = 79.9, MSe = .48, \( p < .001; \ \eta^2_g = .19 \)). There were no significant differences between the ratings of the MCA and MIA (mean MCA = 4.42, mean MIA = 4.39; \( p = .54, \ d = .05 \)). The MCP was rated less interesting than the other two conditions (mean MCP = 3.69; MCA vs. MCP, t (460) = 10.6, \( p < .001, \ d = .99 \); MIA vs. MCP, t (439) = 9.9, \( p < .001, \ d = .96 \)). This confirms choice of the control condition: The meaning incongruent action animation can control for the increased engagement provided by an animation.

Animations were made and displayed using Vizard VR Software (World Viz). The animations were always of the same male avatar acting, on a dark background. Contextual information in the animations were limited to the objects necessary to act out the meaning of the verb (e.g., to chisel, required a chisel as well as an object to chisel — a stone, but the animation contained no other objects).

Note that in the example shown in Fig. 1 there is a spatial separation between the critical object (chisel) and the object that is acted upon. This spatial separation was not intentional and only occurred for 4 out of the 21 MIA animations. In Experiment 2 this spatial separation never occurs.

It may seem counterintuitive to not use videos of actual people or children as this may lead to increased simulation. However, using computer generated animations allowed for very controlled MIA animations, increased consistency between all animations, and (especially in Experiment 2), and the ability to create colorful and attractive material. In addition, animations like this are planned to be used in combination with textbook material being developed by a publisher of school textbooks in the Netherlands.

These short animation videos were embedded within a larger “game” that allowed the child to decide the order that they wanted to see the animations/pictures. This game included three context photographs of a kitchen, a garden and a living room. For each context seven objects were placed around the photograph and the child could click on (single click with the mouse) the object that they wanted to see the animation of. When holding the mouse over an object the corresponding word-to-be-learned would be played auditorily (e.g., they would hear chiseling when hovering over the chisel). Once the child clicked on the object (e.g., chisel) the context photograph would disappear and a close up animation/picture of the relevant action would begin (e.g., chiseling for the meaning congruent animation). The still pictures were displayed for the same amount of time as the animations (4 s). When the animation/picture was finished, the screen would go blank for 500 ms and then the context photograph would reappear. Once all of the objects had been clicked on twice in one context, the next context would be displayed with the seven objects corresponding to the seven words to learn related to that context (e.g., kitchen). After clicking on an object twice and watching the related animation twice, the object would disappear from that context photograph. If they only clicked on an object once, the object would remain until they clicked on the object a second time. Similarly they would not advance to the next context photograph if they had not clicked on all objects twice. This rarely happened, but when it did the experimenter pointed out the object they still needed to click on a second time. The time watching the animations/pictures was held constant for all animations/pictures. In the second session, the same two objects click procedure was used, but the context pictures were presented twice to reach the four total times viewing.

Along with each video animation a sentence using the verb correctly was played auditorily. The sentences were not a dictionary definition of the word, but instead used the verb in a meaningful way (e.g., Bert is een mooi beeld aan het beitelen, word-by-word translation Bert is a beautiful statue PROGRESSIVE chiseling, meaning: Bert is chiseling a beautiful statue.) We chose this for two reasons: First, much of the vocabulary that children learn is presented in exactly this way (see Cain, Oakhill, & Elbro, 2003). Second, by using definitions we would have created a situation where the definition gives necessarily correct information about the meaning. For MIA, the animation would have to give conflicting incorrect information, which could lead to children learning to ignore the animation entirely over the course of the experiment. By using context sentences there is no inherent conflict between what is said in the context sentence and what is shown in the animation. In other words, if you do not know what chiseling means, it allows for the possiblity that the animation is showing something about chiseling. Nonetheless, the context sentence plays an important role for understanding the meaning of the word. We believe it is the pairing of the word and the sentence with the animation that could improve learning, by leading to richer simulation. All sentences corresponded to the format “Subject is adjective object verb-to-learn”. Specifically, the verb-to-learn was always the last word in the sentence. There was always an object of the verb, phrased as an adjective followed by a noun (object). The adjective was used to add contextual meaning to the sentence. On average the sentences were 11.6 words long, (minimum length — 8 words; maximum — 15 words long). The mean length of auditory version of the sentences was 2.8 s (SD = .75 s).

2.1.3. Pretest

In addition to the training videos, we also created a pretest to determine whether the children knew the meaning of our 21 words
to learn prior to watching the training videos. The experimenter read each of the words and recorded the response of the children using a digital recorder. We did not ask the children to give a definition of the word, which is often very hard to verbalize for 7–8 year olds (and even for adults). Instead the children were asked to describe the word in any way they could. Knowledge of a word meaning is not always an all-or-nothing phenomenon: They may know that hoeing has something to do with the garden, but not exactly what hoeing a garden would accomplish. However, to be conservative we decided that if they gave any meaningful reply at all, we scored the word as known. As described below in the Results section, we only look at the training effect for the words that the children did not know previously, based on their answers on this pretest.

2.1.4. Posttest

At the end of the second training session children were tested on their knowledge about the meaning of the 21 words they were asked to learn. This was done with two different measures, a Picture Pointing task and an Object Completion task. The two tasks were given in random order across children.

The rationale for having two posttests was that each tapped into different knowledge about the meaning of the word. In the Picture Pointing task each of the 21 words (plus two practice words) were read aloud by the experimenter one at a time. After the experimenter read the word aloud, the child was asked to point to one of four pictures that fit best with the word. The pictures were full color photographs of objects, displayed on a laptop screen. The correct picture included an object that is relevant for doing the action of the verb. The correct object pictured was never the exact exemplar of the object from the video animations but it was similar to the object pictured in the animation. For example, if the word was to chisel, the correct picture was of a chisel. One of the three foil pictures was a correct picture for a different verb in the same context (for example, for chisel, a hoe was one of the foil pictures). The other two foil pictures were objects that you could clearly perform an action with, but they were not objects that a child would commonly interact with (see Fig. 2 for example).

In the second test of learning, the Object Completion task, children were asked to name an object that is needed in order to perform the action correctly, specifically focusing on the types of objects that are acted upon in this action (for example, What can you chisel? was read aloud by the experimenter and a correct answer would be rock, wood, or any other hard material).

The difference between the posttests can be understood relative to degrees of knowledge of meaning. Lafer and Goldstein (2004) make a distinction between four degrees of knowledge of a word. One distinction is the difference in knowledge about a word between those who can recall the form or the meaning of a word and those who cannot recall but can recognize the form or meaning in a set of options. The second distinction is the difference between being able to provide the word for a given meaning versus supplying the meaning for a given form. Using these two dimensions, the Picture pointing task would be considered to be a passive recognition of meaning (a weak degree of knowledge measure). The Object completion task fits more into the active recall of meaning category. Active recall is considered to be the strongest degree of knowledge of the meaning. The Object completion task fits into a larger category of association vocabulary tests. If a learner is asked about what object is associated with performing a given action (chiseling), the correct answer can only be made if the meaning of the action is known (for a review see Lafer & Goldstein, 2004).

In addition to our experimental materials for the animation training, pretest and posttest, we also administered a number of standardized tests. The intent of these tests was two-fold. First, since a between-subjects design was implemented, we wanted to establish that there were no a priori significant differences in cognitive abilities between the groups of children that were randomly assigned to each condition. Second, we wanted to examine the relationship between the scores on the standardized tests and the scores on our posttests. Initially four different standardized tests were administered. However only two of these standardized tests (Verbal Comprehension subtest of the Wechsler Intelligence Scale for Children (WISC III IQ, Wechsler, 1991, Dutch translated version) and a test of Dutch receptive vocabulary (‘Taaltotoets alle kinderen’, TAK; Verhoeven & Vermeer, 2001) were correlated with our pretest and/or posttest data. For that reason we will keep the discussion of the other tests to a minimum. The other standardized tests administered were the Perceptual Reasoning subsection of the WISC III IQ, (Wechsler, 1991) task), a test of learning style (Visual, Auditory, Kinesthetic learning styles test translated to Dutch, (VAK, Fewings, 2013) and two attention subs tests (Count along subtest and Map searching subtest; Test of Everyday Attention for Children, TEA-Ch: Manly, Robertson, Anderson, & Nimmo-Smith, 1998).

2.1.5. Design

A between-subjects design was implemented, in which all words were presented in the same condition to each child. Each child was randomly assigned to the meaning congruent animation, meaning incongruent animation or meaning congruent picture condition. We chose a between-subjects design to increase the power of the experiment, given that children of this age are usually only asked to learn approximately 20 words (and typically learn less than half of those 20 words: e.g., McGregor, Sheng, & Ball, 2007). Given the practical difficulties of testing numbers of children in school over multiple sessions, it was felt that having the children learn 21 words and making the training condition between subjects, would lead to the best chance of detecting...
differential effects of learning.

2.1.6. Procedure

All children were tested individually over three sessions. Session one began with testing the children on all of the standardized tests (see Materials section above). Additionally at the first session we conducted a pretest of all of the words they would later be asked to learn (see Pretest above). Session one took about 30 min in total. In the second session the child was exposed to the training animation/picture for each word two times, which took approximately 8 min. The children were instructed to watch the animations/pictures and listen to the sentences. They were told that we were evaluating a word learning game to see if it was fun to do. They were also instructed that we would later test them about the words to see how well the game worked. In the final session the child was exposed to the training animation/picture for each word four times, which took approximately 15 min. In total, the children watched the animation/picture six times across the experiment. Following that the Posttest was administered (see Posttest above). The post-test took approximately 15 min to administer. There was one week between training sessions 2 and 3. The time between session 1 and session 2 varied (M = 11 days, minimum = 7 days, maximum = 18 days).

3. Results

We report the results of the standardized tests both overall and by training condition. Inferential statistics are conducted using each of the standardized test scores to test for a priori differences between groups. Additionally a linear regression analysis is conducted to test for the predictive power of the scores on the standardized tests on the Picture pointing test and Object completion test. Following that overall results of the pretest and posttest by condition, and in relationship to the standardized test scores are discussed. Planned comparisons are conducted between MCA and MIA, MCP and MCA as well as MCP and MIA. The rationale for the comparison MCP and MIA was based on the possibility that the MIA condition groups would actually lead to greater learning than the MCP. This could have occurred due to the fact that the MIA condition was rated to be more engaging/interesting than the MCP condition (see ratings test in Methods section). An alpha level of .05 was used for all statistical tests.

3.1. Standardized tests

The mean scores of each of the standardized tests are reported in Table 1.

There were no significant differences between the scores on the TAK across conditions (F (2, 153) = 1.42, MSe = 57.57, p = .25; \( \eta^2_p = .02 \)). Similarly no significant differences were seen in the scores on the Verbal Comprehension subtest of the WISC-III between condition groups (F (2, 153) = 2.23, MSe = 14.86, p = .11; \( \eta^2_p = .03 \)). Likewise, there were no significant differences seen between the scores on the Perceptual Reasoning subtest of the WISCIII (p = .86; \( \eta^2_p = .001 \)), the VAK (p = .38; \( \eta^2_p = .005 \)) or the two subparts of the TEA-Ch (p = .42; \( \eta^2_p = .003 \) for the count along subtest, p = .78; \( \eta^2_p = .005 \) for the map searching subtest) between conditions.

A multiple regression was run to predict the score on the Picture Pointing test from all standardized tests (TAK, Verbal Comprehension subtest and Perceptual Reasoning subsection of the WISCIII, VAK, and the two attention subtests from TEA-Ch). Overall these variables had a small but significant predictive value on the Picture Pointing scores, \( F(8,147) = 2.7, p = .009, R^2 = .13 \). However, it was only the TAK and somewhat the Verbal Comprehension subtest of the WISCIII that added significantly to the prediction (TAK, p = .01, Verbal Comprehension WISCIII, p = .07). None of the other standardized tests added significantly to the prediction. A multiple regression predicting the scores of the Object Completion test from all standardized tests was also significant, \( F(8,147) = 7.0, p < .001 \). It was again only the TAK and the Verbal Comprehension subtest of the WISCIII that significantly contributed to the prediction of the scores (TAK, p < .001, Verbal Comprehension WISCIII, p = .001). There was no significant effect of the other standardized tests; hence we did not include the scores of the other standardized tests in the final MANCOVA. There was a small but significant correlation between the TAK and the Verbal Comprehension subtest, \( (r = .32, n = 156, p < .01 \). In an additional multiple regression analysis including only the TAK and the Verbal Comprehension subtest indicated significant predictive value on the Pretest scores (F (2, 153) = 53.35, p < .001, \( R^2 = .41 \)). However, it was only the TAK that significantly contributed to the prediction (TAK p < .001, Verbal Comprehension WISCIII p = .69).

3.2. Pretest scores

The pretest indicated that on average, children knew 5.3 (SD = 2.2) out of the 21 words prior to participating in the experiment. However there was no significant difference between the number of words children knew across the three conditions (F (2, 153) = 58, MSe = 4.8, p = .56; \( \eta^2_p = .008 \)). The mean words known per condition were as follows: MCA = 5.5, SD = 2.0; MIA = 5.4, SD = 2.2; MCP = 5.0, SD = 2.4).

3.3. Posttest scores

To calculate the amount of words learned, for each participant for each word the result of the pretest was compared to the results for each word on each posttest, a correctly learned score for that word was only given if the child did not know the word on the pretest. Additionally a multiple regression was run to test for the predictive power of the scores on the standardized tests added significantly to the prediction. A multiple regression predicting the scores of the Object Completion test from all standardized tests was also significant, \( F(8,147) = 7.0, p < .001 \). It was again only the TAK and the Verbal Comprehension subtest of the WISCIII that significantly contributed to the prediction of the scores (TAK, p < .001, Verbal Comprehension WISCIII, p = .001). There was no significant effect of the other standardized tests; hence we did not include the scores of the other standardized tests in the final MANCOVA. There was a small but significant correlation between the TAK and the Verbal Comprehension subtest, \( (r = .32, n = 156, p < .01 \). In an additional multiple regression analysis including only the TAK and the Verbal Comprehension subtest indicated significant predictive value on the Pretest scores (F (2, 153) = 53.35, p < .001, \( R^2 = .41 \)). However, it was only the TAK that significantly contributed to the prediction (TAK p < .001, Verbal Comprehension WISCIII p = .69).

Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Overall</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>WISC-III verbal</td>
<td>9.2</td>
<td>3.9</td>
<td>10.1</td>
<td>3.7</td>
<td>8.9</td>
<td>4.2</td>
<td>8.6</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>WISC-III Perceptual reasoning</td>
<td>22.8</td>
<td>11.3</td>
<td>23.3</td>
<td>11.3</td>
<td>23.2</td>
<td>10.2</td>
<td>21.8</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>TAK</td>
<td>80.0</td>
<td>7.6</td>
<td>80.8</td>
<td>6.5</td>
<td>80.5</td>
<td>7.7</td>
<td>78.5</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>VAK-Visual</td>
<td>40.6</td>
<td>17.3</td>
<td>39.9</td>
<td>15.7</td>
<td>41.7</td>
<td>17.0</td>
<td>40.3</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>VAK-Auditory</td>
<td>27.0</td>
<td>13.6</td>
<td>25.0</td>
<td>14.9</td>
<td>26.3</td>
<td>12.9</td>
<td>29.7</td>
<td>13.1</td>
<td></td>
</tr>
<tr>
<td>VAK-Kinesthetic</td>
<td>32.9</td>
<td>15.8</td>
<td>30.0</td>
<td>16.6</td>
<td>32.5</td>
<td>13.8</td>
<td>30.2</td>
<td>16.9</td>
<td></td>
</tr>
<tr>
<td>TEA-Ch — Count along</td>
<td>7.2</td>
<td>2.5</td>
<td>6.8</td>
<td>2.8</td>
<td>7.7</td>
<td>2.1</td>
<td>7.0</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>TEA-Ch — Map searching</td>
<td>17.7</td>
<td>6.6</td>
<td>18.6</td>
<td>6.6</td>
<td>17.3</td>
<td>6.3</td>
<td>17.4</td>
<td>7.1</td>
<td></td>
</tr>
</tbody>
</table>
The mean scores and standard deviations of the Picture Pointing task and the Object Completion task are illustrated in Fig. 3 (unadjusted scores) and Table 2 (adjusted scores).

We conducted a MANCOVA analysis including the three training conditions (MCA, MIA, MCP) as independent variables, the two test types (Picture Pointing, Object Completion) as dependent variables and the scores from two standardized tests as covariates (TAK & Verbal Comprehension subtest of the WISC-III). After controlling for the effect of the covariates, an overall main effect of training condition was found with both the Picture pointing task, $F(2, 151) = 6.01, MSe = 5.39, p = .003$, and the Object Completion task, $F(2, 151) = 5.41, MSe = 3.79, p = .005$. The partial eta squared statistic was .10 for both the Object Completion and Picture Pointing task. The planned comparisons for the Picture Pointing task using adjusted means indicated a significant difference between the MCA and the MIA (Difference = 1.26, SE = .48, $p = .007, d = .54$) as well as between the MCP and the MIA (Difference = 1.44, SE = .46, $p = .002, d = .62$). No significant difference was found between MCA and MCP (Difference = 19, SE = .47, $p = .69, d = .08$).

In the Object Completion task planned comparisons using adjusted means indicated a significant difference between the MCA and the MIA (Difference = 1.24, SE = .43, $p = .002, d = .64$), and between the MCA and the MCP (Difference = .80, SE = .39, $p = .04, d = .41$). No significant difference was found between the MCP and the MIA training conditions (Difference = .44, SE = .38, $p = .25, d = .23$). Because the unadjusted means are close to the adjusted means (see Fig. 3 and Table 2) we re-ran the analysis with the covariates removed: This leads to the same pattern of results and Table 2 above reflects the amount of new words learned, not the absolute number of correct scores on the posttest. The mean scores and standard deviations of the Picture Pointing task and the Object Completion task are illustrated in Fig. 3 (unadjusted scores) and Table 2 (adjusted scores).

The results of Experiment 1 indicate that meaning congruent animations (MCA) lead to a better understanding of the word meaning as measured by the Object completion task. At the same time, both MCA and MCP lead to greater learning compared to MIA as measured by the Picture pointing task.

According to the predictions of the ABL theory (Glenberg & Galisse, 2012), the results on the Object completion task fall out of the MCA condition promoting the child to simulate the action in such a way as to allow for better or more predictions about the words’ meaning. In other words, the richer sensory and motor

### Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Picture pointing</th>
<th>Object completion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Meaning Congruent Animation</td>
<td>10.66</td>
<td>2.34</td>
</tr>
<tr>
<td>Mean Incongruent Animation</td>
<td>9.41</td>
<td>2.33</td>
</tr>
<tr>
<td>Meaning Congruent Picture</td>
<td>10.85</td>
<td>2.34</td>
</tr>
</tbody>
</table>

![Fig. 3. Bar graphs of unadjusted scores on the picture pointing posttest and object completion posttests for Experiment 1.](image-url)
information in a congruent animation, paired with the sentence, makes predictions about things like how a chisel can be used, what would result from chiseling and what types of substances would afford chiseling more likely to be created. According to the ABL theory, these predictions lead to improvement in word learning.

However, the findings do not provide direct evidence for the link between the improved learning and a richer simulation. Experiment 2 is designed to strengthen the argument that the improved learning found in Experiment 1 is due to the predictions that were derived from simulating the action.

5. Experiment 2

In Experiment 2 children learned the meaning of 21 new verbs that were different from the verbs used in Experiment 1. Like Experiment 1, verbs were paired with either a meaning congruent animation (meaning congruent animation, MCA) or a meaning incongruent animation (meaning incongruent animation, MIA).

The main goal was to strengthen the argument that the improved learning found in Experiment 1 is due to the predictions that were derived from simulating the action. To do this we capitalized on a finding within the action understanding research domain and adapted it to our own design. Understanding an action executed by another person has been shown to be sensitive to conceptual level and the perceptual-motor level information (Ondobaka, De Lange, Newman-Norlund, Wiemers, & Bekkering, 2012; Ondobaka & Bekkering, 2013). The conceptual level relates to higher level understanding of the relationship between objects in the world and the perceptual-motor level relates to the relationships between a particular object and the actor. For example, the fact that cups are typically used for drinking coffee is conceptual. The fact you need to bring a cup to your mouth in order to drink the coffee is perceptual-motor. In ABL terms, conceptual level and perceptual-motor level can be mapped onto the predictors and controllers respectively.

Experiment 2 introduces conceptual level versus perceptual-motor level information by manipulating the text that accompanies the animations: The text either highlights the conceptual level of the action first or the perceptual-motor level of the action first. For example, the conceptual level focus sentence for ranging was "Amy arranged the flowers in a vase by putting them in one by one, where the perceptual-motor level sentence for ranging was "Amy put the flowers one by one in the vase in order to arrange them". The information given in the text is thus always the same; it is only the order that is manipulated. Similar text manipulations have been shown to be effective in motor action paradigms, such as speeded grasping (Bub and Masson, 2010; Masson et al., 2013). There the authors used conceptual level sentences like "To contact his mother John used the cell-phone, and perceptual-motor level sentences like "John used the cell-phone to contact his mother". Depending upon whether sentences put initial focus on the conceptual level or perceptual-motor level information, grasping was executed more rapidly (Bub & Masson, 2010; Masson, et al. 2013).

In both Ondobaka et al. (2012) and Bub and Masson (2010), the manipulation of perceptual-motor and conceptual level information is not exactly the same as our implementation. They examined the effect of perceptual-motor and conceptual level information on performing actions. In the current study, the same idea is applied to learning via watching action rather than performing actions. However, the underlying the idea is that there is a hierarchy of information that includes lower sensory level processing (perceptual-motor level) that is related to higher level conceptual processing (our conceptual level; e.g., Clark, 2013; Friston & Kiebel, 2009).

From the ABL theory we predict that the MCA paired with conceptual level highlighted sentence should lead to the greatest learning. The MCA paired with the perceptual-motor level sentence should lead to less, but still an improvement in learning as compared to the two MIA conditions. The MIA paired with perceptual-motor level information should lead to the poorest learning as there is no learning boost from highlighting the conceptual level information or meaning congruent information in the animation.

6. Methods

6.1. Participants

The participants were 114 children in 2nd grade across four different primary schools in Nijmegen and Apeldoorn in the Netherlands. Ten subjects were not included in the final analysis, six due to experiment error (problem displaying animations at time of second session) and four of them were lost due to illness at one session. Of the 104 participants included in the analysis, 52 were female, and the mean age was 7 years, 6 months (range 7–8 years). Seventy-four of the participants were native Dutch speakers. Ten of the participants were bilingual Dutch native speakers, having learned both Dutch and a second language at home. In these cases the second language was typically Moroccan Arabic or Sudanese Arabic. Twenty of the participants were non-native Dutch speakers, having learned Dutch at school. None of the children participated in Experiment 1. The 104 participants were randomly assigned to the conditions, which resulted in the following breakdown: MCA-Conceptual level n = 28, MCA-Perceptual-motor level n = 27, MIA-Conceptual level n = 25, MIA-Perceptual-motor level n = 24.

6.1.1. Sampling

The same sampling technique was the same as in Experiment 1.

6.1.2. Stimulus materials

For this experiment we wanted to scale up the animations to fit the design of additional experiments and other implementations, such as an educational App for tablets. For this reason several changes were made to the original materials. To begin, a new set of 21 verbs was chosen in the exact same manner as in Experiment 1. Linguistic factors, such as word length, number of syllables and the word frequency were kept as similar as possible across words. Using SUBTLEX-NL (Keuleers et al., 2010), the average log frequency was .12 (maximum = 1.25; minimum = 0). The average number of syllables was 2.48 (maximum = 3; minimum = 2). The average length in letters was 7.19 (maximum = 11; minimum = 5). See Appendix A for a list of all verbs used in both experiments.

Like Experiment 1, we created two animations for each verb. However, the animations were improved in terms of their visual quality from Experiment 1. The new animations were slightly longer (6 s compared to 4 s), included either a male or female avatar which appeared more childlike (see Fig. 4 for example of avatar in each condition).


The animations were displayed using Psychopy (version 1.80.03; Peirce, 2007). The procedure for displaying the animations was somewhat different than in Experiment 1, the motivation for this was to allow for comparison with planned future experiments. Here the order of the animations was kept constant (but pseudo-random) across participants. A trial would begin by a green “play button” being displayed in the middle of the computer screen. When a child clicked (either single or double clicked) on the play button, the play
The button would immediately change to yellow. After a 100 ms delay the new word would be played auditorily. After the offset of the auditory word there was a 500 ms delay before the animation started. The related sentence would play starting 100 ms into the start of the animation. All sentences were shorter than the animations. At the end of the animation there would be a blank screen for 1500 ms for session 1 or 300 ms for session 2 before the green play button would appear again and child would have the opportunity to start the next animation. We shortened the time between the end of the animation and the appearance of the green play button in session 2 to reduce the total time of the experiment. Like Experiment 1, the display of the stimuli was self-paced by the children.

Along with each video animation a sentence using the verb correctly was played auditorily. This was the critical manipulation in this experiment. The sentences either had a semantic focus on the conceptual level of the action or on the perceptual-motor level of the action. The semantic focus was manipulated via word order. Sentences with a semantic focus on the conceptual level of the action began by first stating the conceptual level information and then stating the actual movement involved to achieve that action. Sentences with a perceptual-motor level focus reversed this order while keeping the information supplied by the sentence constant (see Table 3 for example stimuli, Appendix A for all stimuli).

On average the sentences were 11.5 words long, (minimum length = 9 words; maximum = 16 words long). The mean length of auditory version of the sentences was 2.8 (SD = .52 s).

6.1.3. Pretest
The same pretest as in Experiment 1 was administered with the new words.

6.1.4. Posttest
The same two posttests as in Experiment 2 were administered with the new words.

6.1.5. Standardized tests
Experiment 1 indicated that only the Verbal Comprehension subtest of the WISCIII (WISCIII IQ, Wechsler, 1991, Dutch translated version) and the TAK were correlated to the scores on our posttest. However, due to time constraints on collecting data in schools and the fact that the TAK appeared to correlate better with both posttests, we decided to only use the TAK standardized test.

6.2. Design
As in Experiment 1, a fully between subjects design was implemented with the factors animation type (meaning congruent or meaning incongruent animation, MCA or MIA) and sentence type (Conceptual level or Perceptual-motor level). This resulted in a $2 \times 2$ design. Each child was randomly assigned to one of the four conditions: meaning congruent animation, Conceptual level sentence (MCA-Conceptual level); meaning congruent animation, Perceptual-motor level sentence (MCA-Perceptual-motor level); meaning incongruent animation, Conceptual-level sentence (MIA-Conceptual level); or meaning incongruent animation, Perceptual-motor level sentence (MIA-Perceptual-motor level).

6.3. Procedure
All children were tested individually over two sessions. Session one began with the TAK and the pretest. Additionally the child was exposed to the training animation/sentence for each word two times. With the exception of the mechanics of how the animations started by clicking on the green play button, instructions about the fact that we were evaluating a word learning game was the same as in Experiment 1. In the second and final session children watched the animation/picture four times, for a total of six times across the experiment. This was the same number of exposures as in Experiment 1. Following that the Posttest was administered. There was one week between the two sessions.

7. Results
The analysis steps were identical to Experiment 1 with the exception of only using scores from the TAK as a covariate (only TAK scores were collected). The posttest scores are analyzed using MANCOVA with the factors animation type (MCA or MIA) and
7.1. Standardized tests

The mean scores of the TAK overall and by condition are reported in Table 4.

There were no significant differences between the scores on the TAK across conditions \( F(3, 100) = .49, MSe = 58.1, p = .69; \gamma^2 = .014 \). The pretest indicated that on average, children knew 4.7 words \((SD = 2.8)\) out of the 21 words prior to participating in the experiment. There was no significant difference between the number of words children knew across the four conditions \( F(3, 100) = .63, MSe = 7.7, p = .63; \gamma^2 = .017 \). The mean words known per condition were as follows: MCA-Conceptual level = 4.6 words, SD = 2.8; MCA-Perceptual-motor level = 4.9 words, SD = 2.5; MIA-Conceptual level = 4.2, SD = 2.9; MIA-Perceptual-motor level = 5.2 words, SD = 2.8.

7.2. Pretest scores

The scores were calculated in the same way as in Experiment 1. The mean scores and standard deviations of the Picture Pointing task and the Object Completion task are illustrated in Fig. 5 (unadjusted scores) and Table 5 (adjusted scores).

We performed the same type of analysis as in Experiment 1, with the independent variables animation type (MCA or MIA) and sentence type (Conceptual level or Perceptual-motor level) on the two test types (Picture Pointing, Object Completion) as dependent variables and the scores from the TAK as a covariate. For the Picture Pointing task an overall main effect was seen for animation type \( F(1, 99) = 29.97, MSe = 5.71, p < .001; \gamma^2 = .23 \), but not for sentence type \( p = .36 \). However, for the Object Completion task a main effect of animation type as well as sentence type was found (animation type, \( F(1, 99) = 29.83, MSe = 6.04, p < .001; \gamma^2 = .23 \) and sentence type, \( F(1, 99) = 14.82, MSe = 6.04, p < .001; \gamma^2 = .13 \). For the Picture Pointing task there was not a significant interaction between animation type and sentence type \( p = .30 \). The Object Completion task also did not indicate a significant interaction between animation type and sentence type, but there was a trend \( p = .08 \).

The planned comparisons for the Picture Pointing task indicated no significant difference between the MCA-Conceptual level and the MCA-Perceptual-motor level, \( \text{Difference} = .93, p = .15 \). There was also no difference between the MIA-Conceptual level and MIA-Perceptual-motor level \( \text{Difference} = .06, p = .93 \).

For the Object Completion task planned comparisons, a significant difference between the MCA-Conceptual level and the MCA-Perceptual-motor level was found \( \text{Difference} = 2.72, SE = .66, p < .001, d = 1.14 \). There was no significant difference in the Object Completion task between the MIA-Conceptual level and MIA-Perceptual-motor conditions \( \text{Difference} = 1.01, p = .16 \).

Unlike Experiment 1, here the covariate TAK was not significantly related to the scores on the Picture Pointing task \( p = .58 \) or the Object Completion task \( p = .18 \). However, we have left the TAK in the analysis since this makes it most comparable to the analysis in Experiment 1 and there is no significant difference in the main effects when it is included in the analysis. Additionally, we have seen in a different follow-up experiment that the TAK was significantly

### Table 3

<table>
<thead>
<tr>
<th>Example sentence stimuli for Experiment 2 by condition.</th>
<th>Dutch sentence plus English gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual level</td>
<td>Amy arranges the flowers in the vase by them (er) one by one into place.</td>
</tr>
<tr>
<td>Perceptual-motor level</td>
<td>Amy places the flowers one by one in the vase to them to arrange.</td>
</tr>
<tr>
<td>Conceptual level</td>
<td>Amy schikt de bloemen in de vaas door ze er een voor een in te steken.</td>
</tr>
<tr>
<td>Perceptual-motor level</td>
<td>Amy steekt de bloemen een voor een in de vaas om ze te schikken.</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Measure</th>
<th>Overall</th>
<th>cCA-Conceptual level</th>
<th>pCA-Perceptual-motor level</th>
<th>cIA-Conceptual level</th>
<th>pIA-Perceptual-motor level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>TAK</td>
<td>80.3</td>
<td>7.6</td>
<td>80.6</td>
<td>7.3</td>
<td>80.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Bar graphs of unadjusted scores on the picture pointing posttest and object completion posttests for Experiment 2.
related to the scores on the posttest, so overall it appears to be an important covariate, even if not for this particular set of data.

Like Experiment 1 a significant difference was found on the pretest between the native and non-native Dutch speakers ($F(1, 102) = 12.03, MSe = 6.84, p < .001, \eta^2 = .11$). The native Dutch speakers knew more of the words-to-be-learned than the non-native Dutch speakers (native Dutch speakers knew 5.3 of our words on average ($SD = 2.8$) where non-native Dutch speakers knew 3.3 of our words on average: $SD = 2.2$). However, there was no significant relationship between native language and the performance on the posttests, Object Completion task ($F(1, 102) = .05, MSe = 8.88 p = .83, \eta^2 = .001$) and the Picture Pointing task ($F(1, 102) = 2.26, MSe = 7.22, p = .14, \eta^2 = .02$).

As in Experiment 1, to verify that there was not an effect of the number of words the children could potentially learn based on the pretest scores, we conducted an additional analysis re-computing the scores on the posttests based on the total number of possible words they could have learned given their pretest score using the same calculation as in Experiment 1. However this lead to the exact same pattern of significant results as with the initial analysis above (Picture pointing task: animation type $p < .001$, sentence type $p = .46$, animation type by sentence type $p = .31$; Object completion task: animation type $p < .001$, sentence type $p < .001$, animation type by sentence type $p = .086$), again with the same pattern of significant planned comparisons.

### 8. Discussion

The current study explored the effect of meaning congruent animations on word learning in children 7–8 years old. In Experiment 1, children were exposed to 21 existing but unfamiliar verbs that were presented with a context sentence paired with a meaning congruent animation (MCA), a meaning congruent picture (MCP) or a meaning incongruent animation (MIA). The key question was whether observing the MCA, by nature of the congruent action being depicted dynamically, leads to better understanding of the words’ meaning than watching a meaning incongruent animation (MIA) or a picture (MCP). Results indicated that both MCA and MCP lead to greater learning as measured by the Picture pointing task. However, our second task (the Object completion task) which was designed to measure a strong degree of knowledge of the meaning showed that MCA led to a better understanding of the word meaning.

In Experiment 2 the key question was whether the improved learning seen in Experiment 1 can be more directly linked to the argument that it is the improved simulation relative to the predictors that enhances the understanding of the word meaning. The hypothesis is that highlighting the conceptual level of an action in the accompanying sentence should lead to a greater understanding of the action observed (Ondobaka et al., 2012), which in turn should allow for better predictions about the meaning of the word (Glenberg & Gallese, 2012). This should be particularly evident when learning is measured by the Object completion task. The results concur with this hypothesis. When children saw a meaning congruent animation paired with a sentence that highlighted the conceptual level information, they learned significantly more words as measured by the Object completion task compared to when the animation was paired with perceptual-motor level sentence. No difference was seen on the Picture pointing task between the two sentence types. This is likely due to the fact that this task reflects passive recognition knowledge about the word, and therefore does not reflect the benefit that the MCA-conceptual level sentences added to their learning. Note we did not predict that an effect of conceptual level sentence would actually be seen in the Picture pointing task for this same reason.

There are several alternative interpretations and limitations of the current study. With regard to the results of Experiment 2, our sentence focus manipulation is confounded by word order. The conceptual-level versus perceptual-motor level sentences were constructed by manipulating sentence order, this resulted in the conceptual-level sentences beginning with the word to learn, where the perceptual-motor level sentence had the word to learn at the end. Possibly this order difference lead to an effect of primacy for the conceptual-level sentences, while the perceptual-motor level sentences suffered from cognitive load. While we cannot fully rule out this possibility, we find both the primacy argument and the cognitive load argument unlikely explanations for several reasons. First and most importantly, regardless of the sentence condition, the target word is always presented before the sentence is presented. This gives the opportunity to connect the word to the animation right from the beginning, regardless of the sentence condition. Secondly, effects related to primacy are typically seen relative to recalling or recognizing the same word (e.g., recalling chisel in “Bert is chiseling…”; Deese & Kaufman, 1957). In our posttests they never need to recall the actual word (chiseling), but instead are given the word chiseling and they need to recall information related to the meaning of the word. Additionally, if the effect with conceptual-level sentences was due to primacy, then the perceptual-motor level sentences should benefit from a recency effect, as recency effects are known to be at least as strong, and sometimes stronger than primacy effects (e.g., Morrison, Conway & Chein, 2014). Yet the benefit of recency is not seen for the perceptual-motor level sentences. Furthermore, if the sentence manipulation primarily reflected primacy effects, larger differences should be found in the Picture pointing test as well as the Object completion test, but in fact no differences are seen in the Picture pointing test. Finally, in relationship to cognitive load, if children exposed to perceptual-motor level sentences were having difficulties due to cognitive load, difficulties should be seen in the Picture pointing test, as is seen for example in the MIA in Experiment 1. However, the perceptual-motor level condition children perform just as well as the conceptual-level condition.

Another possible explanation of our results is simply that children are able to learn more words when the words are coupled to meanings that children are already familiar with, for example pairing a word to learn with an MCA or MCP. However, we would argue that it is not the familiarity of the movements per se that improves learning. In Experiment 2 watching familiar actions (MCA) paired with the perceptual-motor level sentences did not lead to greater learning than watching unfamiliar actions (MIA), as is illustrated in Fig. 5 (comparing MCA-perceptual-motor level sentence to MIA-conceptual level sentences). This would not be expected if the improved learning was due to simple familiarity of the meaning.

A further alternative interpretation of the results is in relationship to the type of information that is needed to perform well on the posttests. Since the Picture pointing task required children to choose a picture of an object that is relevant for doing the action of the verb, where the Object completion task involved them naming...
an object that is needed to perform the action correctly (focusing on the objects that are acted upon), it is possible that the difference between the two tests is not based on degrees of knowledge, more passive or active knowledge, but instead is based on the aspect of the action that is being recalled. Our data cannot currently rule out that possibility. To rule out this possibility we would have to re-run this experiment changing the objects in the Picture pointing task to represent the objects that are acted upon and change the Object completion task to focus on the objects that are relevant for doing the action of the verb. However, a priori there is no reason to expect that objects that are acted upon are more difficult to learn or remember than the objects that are used to perform a given action. Where there is an extensive literature about the difference in difficulty between recognition and recall, or active versus passive knowledge (see Yonelinas, 2002 for a review). It seems more likely that the differences between the two posttests are due to degrees of knowledge rather than memory for the object specific aspect of the action.

Overall the results fit well with the idea that representational animations improve learning (e.g., Höfler & Leutner, 2007; Brucker et al. 2015). Furthermore, while that previous research suggests that learning procedural motor knowledge benefits most from using animations; our research extends this to learning verb meanings, at least verb meanings that denote biologically manipulative actions. It is an open question whether animations illustrating the meaning of a non-biological action would similarly lead to improved learning. However, if our results are due to improved simulation from watching a person make the meaning relevant movement, then we would speculate that animations of non-biological actions would not lead to the same benefit in learning. A similar prediction has been suggested elsewhere (Castro-Alonso et al. 2014).

It should be noted that the overall scores in Experiment 2 were higher than the scores in Experiment 1. However, we cannot say why these differences occurred. The experiments involved different words, different children, different animations and different sentence structures, so there are many possible contributions to this difference.

An interesting question regarding this research is what do children in the MIA condition actually learn? These children hear the word used correctly in a sentence and they do see all objects that are needed in order to perform that action, but they see the wrong action being performed. We have no doubt that when a hoeing movement is done with a chisel in the person’s hand, accompanied by a sentence saying that “Bert is chiseling a beautiful statue”, children may well get confused about what chiseling means. Technically speaking, if they focus on just the sentence and the object shown, they could still perform better than chance on our Picture pointing posttest, since they only need to recognize that chisel was something that was shown with the word chiseling, without really knowing what the meaning of chiseling is. In fact they do perform above chance on the Picture pointing task, suggesting that they are getting at least this amount of information out of the meaning incongruent animations.

A priori we thought that the MIA-conceptual-level sentence might lead to slightly better word learning compared to the MIA-Perceptual-motor level condition. The fact that there is no effect of the conceptual-level sentence in the Experiment 2 MIA conditions is easily accounted for: They were not able to map the conceptual-level information onto the animation they see. However, the inclusion of the conceptual plus perceptual-motor level information in all the sentences makes the sentences closer to an actual definition of the word, which may in part be why they perform better overall in Experiment 2.

Together the results generally support embodied theories of word meaning (e.g., Barsalou, 1999) and in particular the ABL theory (Glenberg & Galleeze, 2012) of word learning, as applied to the domain of learning verbs. The results suggest that although sentences about an action and pictures of an action provide some context for the child to simulate the meaning of the word, meaning congruent animations lead to simulations of the action that allow the child to make more accurate/comprehensive predictions about the meaning of the word.

A question one might pose about the simulation explanation of the current data is how can children start a simulation, or predict an action on the basis of meaning congruent animations and conceptual level information when they do not know the meaning of the verb? It has been argued elsewhere that it is necessary to posit representations of action whose content is non-conceptual (e.g., Pacherie, 2011). The relevant point being that perceptual experience can be represented without an understanding of the concepts underlying that perception, and in fact is suggested to be an explanation for the acquisition of observational concepts (see Pacherie, 2011). Furthermore, the context sentence plays an important role in supporting the simulation. It is the pairing of the context sentence to the animation that leads to improved learning. The results of Experiment 2 further support the idea that the context sentence is important for understanding the meaning that is illustrated by the animation.

In general, our results also fit with and extend more traditional models of word learning, such as the Dual Coding theory (Paivio, 1986; Sadoski, 2005, 2009). The Dual Coding theory assumes that in order for word learning to be effective, learners need to build connections between verbal and nonverbal mental representations of the same item in order to learn that item. More recently, this model has also been cast as an embodied model, suggesting that the verbal code deals with language in several sensory modalities and the nonverbal code deals with images in all sensory modalities (see Sadoski, 2009). Although Dual Coding does not offer a specific prediction in relationship to word learning via animations, our meaning congruent animations, pictures and sentences would generally be predicted to help build connections between images (in multiple sensory modalities) and the verbal code.

An important aspect of the results we found is the difference between the scores on the two posttests. Meaning congruent animations had the largest effect on learning as measured with the Object completion task. A result that became stronger when the accompanying sentence focused on conceptual level information. The two posttests were designed to tap into different levels of understanding of the word meanings and they indeed lead to very different results.

Overall, children performed much better on the Picture Pointing task than they did on the Object Completion task. However, we chose the Object Completion task to measure an extended understanding of the word meaning. Naturally, children do not acquire this level of understanding for all of the words that they were trained on over the relatively short course of this experiment, so the lower scores on the Object Completion task actually confirms that it captures an advanced result.

Scores on the Object Completion task, were quite low in Experiment 1 ($M = 3.94$ in MCA, range $1-9$ words, $d = 54$), but were twice as high in Experiment 2 ($M = 8.96$ in MCA-Conceptual level, range $3-14$ words, $d = 114$). Recall that children saw six repetitions of a 4–6 s presentation for each word, for a total of 24 s (Exp 1) or 36 s (Exp 2) per word. We assume that if children were exposed to these animations as part of a larger unit of language instruction (as they would in a realistic classroom setting), there would be more exposures to the word over a longer period of time and the exposures would fit within the theme of that unit, the scores would likely go up even more.

For the same time, learning on average 4 to 9 words well enough to actively recall information about the meaning of the word, after having been exposed to that word only six times is promising. This type of vocabulary instruction could be implemented in the classroom as part of a larger unit including the new words to learn.
Although for this experiment we specifically decided to not actually provide the definition of the word, in practice providing the definition along with a sentence using the word and the meaning congruent animation could be made for teachers to use with their lessons, or for additional study at home. Similar such animation material is already being offered by the largest school textbook publisher in the Netherlands.

8.1. Learning words from other word classes

The purpose of the current study was to explore whether meaning congruent animations would aid word learning more than static pictures and to explore the extent to which the learning words with animation is sensitive to properties related to the process of understanding an action. However, the current study only investigated verb learning. Since this was a new area of research, we chose to focus on verbs: If meaning congruent animations showing motor actions could be effective for word learning, this effect should certainly be observable with verbs. The ABL theory of word learning (Glenberg & Gallese, 2012) suggests that the benefit of animations should also hold for other classes of words. Although the ABL theory focuses on the relationship between language and action, they do not claim that all language phenomena can be accommodated by the action systems. They point out that there is also strong evidence from the embodied cognition literature that the perceptual systems and emotional systems also contribute to language comprehension. That said, the ABL theory still outlines how nouns and determiners can be learned via controllers and predictors. For example, when seeing a particular object, such as a cup, attention is called to the object named by the noun (cup). Typically this is realized as a saccade of the eyes to the location of the object. Secondly, the attention to the object makes available the affordances of the object. The motor realization of the affordances is proposed to occur via activation of mirror neurons and canonical neurons (see Glenberg & Gallese, 2012). In this case, an animation of the object (noun) to be learned should be more effective for learning that noun than a picture, but actually interacting with the real object would be most beneficial. The reason is that a picture of an object does not afford the same things the actual object (see Snow, Pettypiece, McAdam, McLean, Stroman, Goodale & Culham, 2011). Presumably, watching an animation of an object being used correctly, while not the same as actually interacting with the object, would be a close second.

9. Conclusion

In an innovative study, we show that meaning congruent animations paired with sentences can have an effect on word learning that is as strong as the effect of showing pictures when using a recognition task, but the effect of animations is superior to that of pictures when using a task that requires active recall of meaning related information. Secondly, learning words from meaning congruent animations is sensitive to conceptual level properties related to the understanding an action. Applying findings from the action understanding domain to the design of meaning congruent animations can lead to more effective word learning materials, at least for learning verbs. With recent advances in creating and displaying animations, meaning related animations should be encouraged to be used in classrooms to increase the quality and effectiveness of learning.

Acknowledgments

This research was supported by a grant from the National Initiative Brain & Cognition program, #056-33-016. We would like to thank Elise Brummer, Merel Fokkema, Ingeborg Morelissen, Janneke van Druenen and Ceci Verbaarschot for help in collecting the data as well as all of the schools and children that participated in these studies.

Appendix A

Words and sentences (and English gloss) for Experiment 1 & 2.

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>beitiesen</td>
<td>Bert is een mooi beeld aan het beitiesen.</td>
</tr>
<tr>
<td>to chisel</td>
<td>Bert is a beautiful sculpture PROGRESSIVE chiseling.</td>
</tr>
<tr>
<td>boesteren</td>
<td>Bert gaat van rode klei een mooi gezicht boesteren.</td>
</tr>
<tr>
<td>to mold/sculpt</td>
<td>Bert will of red clay a nice face sculpt.</td>
</tr>
<tr>
<td>conserveen</td>
<td>Bert doet lekker kersen in een pot om ze te conserveen.</td>
</tr>
<tr>
<td>to conserve (jam)</td>
<td>Bert puts nice cherries in a jar to them to preserve.</td>
</tr>
<tr>
<td>dartelen</td>
<td>Bert en Mathijs spelen leuk samen; nu zijn ze door de tuin aan het dartelen.</td>
</tr>
<tr>
<td>to frolic/romp</td>
<td>Bert and Mathijs play nice together; now are they in the garden PROGRESSIVE frolic.</td>
</tr>
<tr>
<td>dobelen</td>
<td>Bert en zijn vrienden doen een leuk spelletje, ze zijn aan het dobelen.</td>
</tr>
<tr>
<td>to throw dice</td>
<td>Bert en his friends play a nice game, they are PROGRESSIVE throw-dice.</td>
</tr>
<tr>
<td>dompelen</td>
<td>Voor hij de grote bomen plant, wil Bert ze even in water dompelen.</td>
</tr>
<tr>
<td>to plunge/emerge</td>
<td>Before he the big tree plants, wants-to Bert them shortly in water plunge.</td>
</tr>
<tr>
<td>draperen</td>
<td>Bert helpt om de mooie slingers over de kerstboom te draperen.</td>
</tr>
<tr>
<td>flammenen</td>
<td>Bert gaat de mooie slangenkoeken in een pan flamberen.</td>
</tr>
<tr>
<td>to flambé</td>
<td>Bert will the big pancakes in a pan flambé.</td>
</tr>
<tr>
<td>garneren</td>
<td>De mooie taart is klaar en Bert gaat ze met slagroom garneren.</td>
</tr>
<tr>
<td>to garnish</td>
<td>The pretty cake is ready and Bert will-soon now with whipping-cream garnish.</td>
</tr>
<tr>
<td>harken</td>
<td>Het gras ligt vol met oude bladeren; Bert helpt ze bij elkaar te harken.</td>
</tr>
<tr>
<td>to rake</td>
<td>The lawn is full with old leaves; Bert helps them together to rake.</td>
</tr>
<tr>
<td>ontrafelen</td>
<td>De hele computer kabel zit in de knoop en Bert is het aan het ontrafelen.</td>
</tr>
<tr>
<td>to unravel</td>
<td>The all-of computer cable is in a tangle and Bert is PROGRESSIVE unravel.</td>
</tr>
<tr>
<td>plamuren</td>
<td>De gaten in de oude muur moeten gedicht worden en Bert gaat nu plamuren.</td>
</tr>
<tr>
<td>to fill up/putty</td>
<td>The holes in the old wall have-to filled-up be and Bert will-soon now putty.</td>
</tr>
<tr>
<td>plooien</td>
<td>Voor een nieuw rolje is Bert de stof aan het plooien.</td>
</tr>
<tr>
<td>to fold/crimp</td>
<td>For a new skirt is Bert the fabric PROGRESSIVE crimp.</td>
</tr>
<tr>
<td>bert maakt vers eten voor de baby klaar; hij moet nu de wortels purren.</td>
<td></td>
</tr>
</tbody>
</table>
Amy maalt de pitten door ze eerst te stampt de pitten. Amy cultivate little-tomatoes by them in a greenhouse let to grow. Amy kweekt tomaatjes door ze in een kas te laten groeien. Bas scoops the water from his boat by it (er) with a bucket out-of transfer-by-bucket. Bas hoost het water uit zijn boot door het er met een emmer uit te scheppen. Amy repairs the airplane by the wings to-it glue. Amy herstelt het vliegtuig door de vleugels eraan te lijmen. Bas buckle him-self (er) to (her) with a seat-belt. Bas gespt zichzelf vast door de riem vast te maken. Amy feels one's way to her tru frunnik a sweater when she something nervous does. Amy voelt rond met haar handen als ze naar de lichtknop tast. Bas forces the door by the screwdriver (er) in to put. Bas forceert de deur door de schroevendraaier er in te steken. Bas prunes the plant by (er) branches from to remove. Bas knipt takken af om de plant te snoeien. Bas solders the electrical wires by them to each-other. Bas soldeert de elektrische draadjes door ze aan elkaar vast te smelten.
verjagen
to chase away

Amy feels-her-way to the light-switch as she with her hands around grapples.
Bas slaat naar de vlieg om het te verjagen.
Bas verjaagt de vlieg door er naar te slaan.
Bas chases-away the fly by it to-direction to swat-at.

References


Barba, J. (2013). Nonlinguistic aids in second language text with multimedia aids and the in

Baltova, I. (1994). The impact of video on the comprehension skills of core French


Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehensi


