Robots Learning to Say "No": Prohibition and Rejective Mechanisms in Acquisition of Linguistic Negation

FRANK FÖRSTER and JOE SAUNDERS, University of Hertfordshire, UK
HAGEN LEHMANN, Università di Macerata, Italy and University of Hertfordshire, UK
CHRYSTOPHER L. NEHANIV, University of Hertfordshire, UK and University of Waterloo, Canada

"No" is one of the first ten words used by children and embodies the first form of linguistic negation. Despite its early occurrence, the details of its acquisition remain largely unknown. The circumstance that "no" cannot be construed as a label for perceptible objects or events puts it outside the scope of most modern accounts of language acquisition. Moreover, most symbol grounding architectures will struggle to ground the word due to its non-referential character. The presented work extends symbol grounding to encompass affect and motivation. In a study involving the child-like robot iCub, we attempt to illuminate the acquisition process of negation words. The robot is deployed in speech-wise unconstrained interaction with participants acting as its language teachers. The results corroborate the hypothesis that affect or volition plays a pivotal role in the acquisition process. Negation words are prosodically salient within prohibitive utterances and negative intent interpretations such that they can be easily isolated from the teacher's speech signal. These words subsequently may be grounded in negative affective states. However, observations of the nature of prohibition and the temporal relationships between its linguistic and extra-linguistic components raise questions over the suitability of Hebbian-type algorithms for certain types of language grounding.

CCS Concepts: • **Computing methodologies → Discourse, dialogue and pragmatics**; **Cognitive robotics**; **Cooperation and coordination**; *Cognitive science*;

Additional Key Words and Phrases: Developmental robotics, language acquisition, symbol grounding, human-robot interaction

ACM Reference format:
Frank Förster, Joe Saunders, Hagen Lehmann, and Chrystopher L. Nehaniv. 2019. Robots Learning to Say "No": Prohibition and Rejective Mechanisms in Acquisition of Linguistic Negation. *ACM Trans. Hum.-Robot Interact.* 8, 4, Article 23 (November 2019), 26 pages.
https://doi.org/10.1145/3359618

23

The work was supported by the EU Integrated Project ITALK ("Integration and Transfer of Action and Language in Robots") funded by the European Commission under contract number FP-7-214668.
Authors' addresses: F. Förster and J. Saunders, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK; emails: {f.foerster, j.1.saunders}@herts.ac.uk; H. Lehmann, Università di Macerata, P.le Luigi Bertelli, 1, Macerata, 62100, Italy, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK; email: hagen.lehmann@unimc.it; C. L. Nehaniv, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK, University of Waterloo, 200 University Avenue W., Waterloo, N2L 3G1, Canada; email: cehaniv@uwaterloo.ca.

2573-9522/2019/11-ART23
https://doi.org/10.1145/3359618

1 INTRODUCTION

In research on early language, we often find the claim that children's early productive vocabularies were dominated by nouns referring to concrete objects such as foods or toys. This assumption appears to have been picked up and reinforced by work in robotics on symbol grounding (cf. Stramandinoli et al. [48]). As a consequence there are plenty of studies that focus on the acquisition of precisely these types of words [20].

The importance of mutual or joint reference between mothers and children to perceptible objects and events is emphasized by more recent, so called usage-based theories of language development [51]. Mother, child, and external referent make up a triadic joint-attentional frame that is very much the focus of these later theories. Cognitively such triadic interactional constellations are more complex than a simple dyadic interaction.

In those areas of developmental robotics concerned with language acquisition in artificial agents, the linguistic units in focus are similarly those that can be construed as referents for concrete physical objects, object properties, or perceptible events. Central in this area of research is the notion of symbol grounding, the construction of links between abstract symbols and signals or constructs that are based on the embodiment of the agent [18]. The construction of such links may be regarded as a form of sense making with respect to the linguistic entities under consideration. The linguistic units in question are typically words or simple grammatical constructions. The agent's embodiment often presents itself with respect to a stream of sensorimotor data.

In stark contrast to this focus on concrete referents, and the words that label them, is the observation that amongst the very first words produced by a toddler are many that do not fall into this category. We find words such as "no," "hi," or "bye" [11], which are sometimes referred to as *social words*. Negation words are thus amongst the very first words in many toddlers' active vocabularies and are used by them to reject things or to self-prohibit [17, 53], the latter being a function that is rarely seen with adult speakers. The idea to link these social words with a robot's sensorimotor data appears strangely inappropriate. The question then is how these types of words should be handled in an embodied language acquisition framework. In the following, we will provide one possible answer to this question.

1.1 Overview

1.1.1 Location of Work. The presented work pulls together several strands and areas of research to answer the questions where the earliest forms of linguistic negation may originate from, and how they may be acquired by a linguistically incompetent learner. The theoretical motivation and background are mainly located in developmental psychology and psycholinguistics, with the most important author being Roy D. Pea who wrote his dissertation on this subject [32] (summarized in Pea [33]).

On the methodological level, and given that our experiments crucially involve dyads consisting of humans and a small humanoid robot, the work falls into the area of Human-Robot Interaction (*HRI*). In no less important ways it builds up upon recent research on symbol grounding in developmental and cognitive robotics [7, 34, 35, 39, 40]. We adopt a constructive, or synthetic, approach to cognitive science that has been termed 'cognitive developmental robotics' elsewhere [2]. Under this approach robots are used to elucidate naturally occuring cognitive processes and phenomena. Additionally, we combine this approach with linguistically unrestricted human-robot interaction as language acquisition is an inherently social and interpersonal process. Taking the social aspect of this process seriously has direct methodological consequences. Social processes are necessarily distributed across two or more agents. This means that the appropriate unit of analysis is the dyad or group, rather than the individual. In terms of machine learning this implies, that we need to widen our analytical scope such that it extends beyond the individual learner's brain. This brings attributes and phenomena into scope that would be overlooked otherwise. One such attribute is the timing of linguistic productions that, as we will show later, has important consequences for the success rate of the acquisition or learning algorithm.

The main experiment on the role of prohibition in negation acquisition described in the present publication was designed in tandem with an experiment on so called intent interpretations described in [14]. The two experiments were jointly designed to achieve maximum alignment for the purpose of comparability. We will therefore frequently refer to the latter as *rejection experiment* and will use it as comparative "benchmark". Both negation experiments crucially involve affect or motivation, and both build up on "non-affective" work on robotic language acquisition conducted by Saunders et al. [38–40]. The experimental setup of both negation experiments and the work of Saunders et al. are similar to such a degree that they allow us to conduct comparisons between participants' speech between these experiments. The most important difference between the two negation and *Saunders' experiment* is that the latter did not model affect. Contrary to the two negation experiments, *Saunders' experiment* was genuinely concerned with the acquisition of referentially concrete nouns, verbs, and adjectives. However, all three experiments utilized the same robot and took place in the same room with the same physical setup to render the results comparable.

In terms of its location within HRI, the present work might be considered unusual in that we use HRI as a methodological tool to answer questions that, in principle, lay outside of the field. Rather than attempting to come up with design principles for robots or robotic behavior, or trying to shed light on human preferences with respect to robots, with the aim of improving human-robot interaction, we use a combination of HRI and developmental robotics as testbed for developmental theories of language acquisition. As is customary for constructive approaches, we simultaneously try to gain insights that help to improve the design of cognitive architectures and inform socially oriented machine learning.

1.1.2 Research Aim and Research Questions. The overarching aim of both the prohibition experiment, presented in the present publication, and the rejection experiment, presented elsewhere [14], is to understand the developmental origin of linguistic negation. Where do the first actively used negation words of human toddlers, chiefly "no," originate, and how do toddlers learn to use them? Given that we use a constructive methodology, these questions more or less equate to the question: Based on our current understanding of the developmental origins of negation, are we able to build a machine that is able to acquire negation words in a similar fashion as humans do? More specifically, and given the focus of this work on Spitz' hypothesis described below (Section 2), this translates to the following research questions:

Can we replicate the interpersonal mechanism proposed by Spitz' on the origin of linguistic negation in silico, that is, by having human-robot dyads as "stand-ins" for parent-child ones?

Granted that this replication involves a number of processes and components, this question breaks down into the research questions *Q2* to *Q5* below. Question 1 is more general in that it forms the basis of both negation experiments and is not specific in terms of the concrete hypothesized mechanism.

(Q1) Can we corroborate the psycholinguistic hypothesis that affect or motivation are necessary requirements for the acquisition of linguistic negation using a constructive or synthetic approach involving human-robot dyads?

(Q2) Do human participants, when confronted with a defiant humanoid robot, and in a situation where the robot tries to touch forbidden objects, produce linguistic prohibitions?
(Q3) If yes, are these prohibitions good sources of negation words, that is, do they contain negation words, and if yes, are these prosodically salient?
(Q4) If yes, how do prohibitions compare to negative intent interpretations both in terms of the frequency of the contained negation words as well as their prosodic saliency rates?
(Q5) If all of the above holds, then will the felicity rate of the robot's active uses of negation words improve in response to the more frequent presence of negation words and relative to the baseline established by the rejection experiment?

1.1.3 Overview of Paper. In the remaining part of the Introduction, we will point out the contributions made in the presented work and introduce the notion of symbol grounding. Symbol grounding plays a central role in the language acquisition architecture described in Section 3. Moreover, symbol grounding may constitute the main difference between embodied approaches to language acquisition, and disembodied approaches that are more common in natural language processing.

Section 2 highlights the role of affect or motivation in early language acquisition, describes the rationale behind the two negation experiments, and gives a high-level description how we operationalize motivation or affect in both the experimental setup and the cognitive architecture.

Section 3 describes the study setup and sketches the cognitive architecture that formed the basis of both the robot behavior as well as the learning components involved in the acquisition process on the robot's side.

Section 4 describes the methodology for analysing the negative speech produced by both human participants and robot within the experiment. A central part of the analysis form two functional or speech act taxonomies that were developed to categorize the negative utterances that we observed in the experiments. The first taxonomy extends Pea's taxonomy of negative child utterances, the second one was developed as "symmetric" counterpart to the first taxonomy to categorize participants' negative speech. This section also translates the research questions from the previous section into hypothesis formulated in terms of the experimental variables.

Section 5 describes the results of the analysis conducted on the gathered speech and interaction data. Participants' negative speech is analysed on three levels: utterances, corpus (or word), and functional level. This allows use to link word frequencies to speech act or type frequencies and draw conclusions as to where the negative words in the robots lexicon originated from. The robot's negative speech that it eventually produced is judged both for its adequacy or felicity in the respective conversational and interactional context as well as for its functional types. Felicity rates of the robot's speech, as judged by external coders, are subsequently statistically compared between the two negation experiments.

Section 6 finally discusses the findings of the experiment.

1.2 Contributions

The present work makes the following contributions:

- We present the first example of a cognitive architecture that operationalizes affect and motivation for the purpose of symbol grounding.
- The negation experiments described herein provide a first example of a robot that acquires negation words such as "no" from the unconstrained speech of a human tutor
- We provide constructive quantitative support for the hypothesis that parental prohibition may constitute (a part of) the developmental origin of human linguistic negation (*Spitz' hypothesis*)

- Under the assumption that *Spitz' hypothesis* is true, our experimental data provides support to counter the hypothesis that the core learning mechanism underlying word learning is mainly a process of association (Hebbian learning in the wider sense). Given the stated assumption, more powerful core learning mechanisms are likely to be required, with reinforcement learning being a likely candidate.

1.3 Symbol Grounding

Artificial symbol grounding, or perceptual symbol systems [4], attempt to solve or break out of the symbol grounding problem, the formulation of which is frequently attributed to Harnad [18]: If symbols are recursively defined or explained merely by the concatenation of other symbols, as is the case in a dictionary, then how can the agent make sense of such symbols given the often circular relationship between the explanandum and explanans (see Reference [34] for an example)? The principle method employed in solving this problem is to connect some or all symbols of the system to sensorimotor data that originates in the agent's own embodiment: its visual sensors, its haptic sensors, its auditory perception, or any kind of derivative constructs that are computed from data originating from such sensory channels. The way existing symbol grounding systems differ from each other is mainly in the method how the link between symbolic and sensor-derived data is established. The methods for constructing and maintaining such links may involve neural networks [7, 49], symbolic artificial intelligence approaches [10, 42, 46], statistical learning methods [47], or methods inspired by an enactive approach [25, 26]. The latter are typically data-driven and, arguably, representation-free in the sense that no models of object or event categories are constructed. Technically, this is made possible through the use of lazy learning algorithms [1], which compute retrieval or classification requests directly on the "remembered" data.

The notion of symbol grounding is ambiguous in more than one way. On one hand, it is used both in a static and dynamic way. In the dynamic variant, authors use the word to describe the wider processes through which symbolic and sensor data come to be associated with each other. In the static variant the focus is more narrowly on the details of the relationship between the "groundee", typically the sensorimotor data, and the grounded, the symbol. On the other hand, symbol grounding has also been used in the context of "common ground" in the modelling of dialogues. We refer to Kennington et al. [23], who briefly discuss the differences and for further pointers into the literature. In the following, we will use symbol grounding in the first sense of the word, and mostly in its static meaning.

As mentioned in the introduction, most of the existing work on symbol grounding focuses on and is limited to the grounding of words that may be seen to either refer to concrete objects ("concrete nouns") or to temporally unfolding perceptible events and processes ("concrete verbs"). More recently Stramandinoli et al. [48] provided an example for grounding more abstract verbs such as "use" or "make" on the back of already grounded concrete verbs such as "cut" or "slide." While this is certainly an improvement in terms of the ability to ground a more general class of words, it is not clear how this approach would help to ground social or socio-pragmatic words such as "no," "yes," or "hi," or emotion words such as "sad" or "upset." The present work is therefore not an argument in favour or against any of the particular established symbol grounding methods. More generally, it is an attempt at widening the scope of what symbols can be grounded in. Instead of solely grounding words in sensorimotor data, or derivatives thereof, we advocate in favour of widening the scope of potential grounding targets to encompass affect or motivation.

2 NEGATION AND AFFECT IN LANGUAGE ACQUISITION

Authors such as Pea emphasize the significance of affect in the context of acquisition of early linguistic negation [33]. Less well understood are the concrete ramifications of this primacy of affect for a cognitive architecture in terms of required components, learning mechanisms, or the dynamics of the learning process. Hence roboticists have so far been unable to create machines that could acquire and engage in this aspect of human speech.

2.1 Spitz' Hypothesis

Spitz' hypothesis [44], as summarized in Pea [33], asserts that infants' major source of negation words is rooted in parental prohibitive utterances that occur in conjunction with the child's "incomplete act." The latter is thought to cause a "frustration of the child's id drives" and negative affect. The child is subsequently thought to associate this negative "affective cathexis" with the negative word or gesture. Via role reversal, or, in Pea's words, "the identification with the prohibiting parent" the child is then thought to use these negative symbols for the purpose of rejection. It is important to notice that this hypothesis, as summarised by Pea, does not specify what precisely causes the child's act to be incomplete. However, the tacit assumption seems to be that the parent has some non-trivial role in this. Technically, what or who precisely is the cause of frustration is an important point, as we will discuss in Section 6.

2.2 Intent Interpretations

An alternative, yet mutually non-exclusive, hypothesis on the origin of children's first negation words involves the notion of intent interpretations. Intent interpretations are linguistic descriptions or ascriptions of the addressee's affective or volitional state. Ryan hypothesized that they played a vital role for infants to learn how to express their intent [37]. Pea subsequently observed that such intent interpretations often contain lexical negatives when the target expressions involve negative affect or motivation [33]. Based on this observation, we hypothesized that precisely these intent interpretations may constitute an alternative source for the learner's first negation words. The mechanism here is simpler than in the case of prohibition in that it simply requires the regular temporal co-occurrence of negative affect and negative word. Role reversal or identification with the parent are not needed.

2.3 Rejection Experiment

In the *rejection experiment* [14], we tested whether the robot's display of affect would trigger intent interpretations on part of our participants akin to what Pea and Ryan had observed in parent-child dyads. Given our focus on negation, we were particularly interested in the negative variant of intent interpretations, that is intent interpretatios with respect to displays of negative affect or motivation. As described in Förster et al. [14] the effect of the robot's affective displays on participants' speech was remarkable, and did confirm the potential of intent interpretations as source for negation words. We use the results of the *rejection experiment* within the present article as point of reference to assess the comparative efficacy of the prohibition task in terms of the acquisition of negation words. The language acquisition architecture employed in the present study is identical to the one used in the rejection experiment. For the learning to be efficacious, the system relies on strong interactional regularities that operate on relatively small time windows of several hundred milliseconds to very few seconds. More simply expressed, the human parent, teacher, or tutor needs to express the relevant word roughly at the time when the respective referent or affective expression is present. We will refer to this as the *simultaneity constraint*. In the paradigmatic joint attentional frames, that feature centrally in modern developmental account of language acquisition, the simultaneity contraint can be assumed to be fulfilled by default: a triadic joint attentional frame is only given, when all three components of the triad are temporally co-present: tutor, learner, and referent. In Saunders' experiments, which focussed on the acquisition of object labels and properties, the simultaneity constraint was typically also adhered to by the

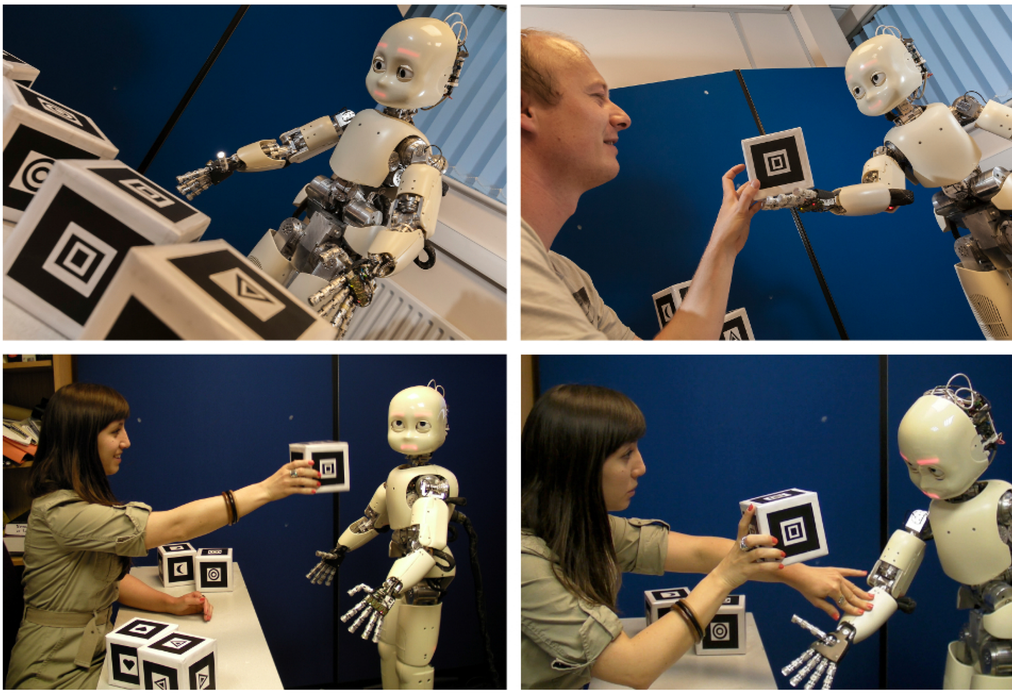


Fig. 1. Experimental setup and interactive robot behaviors as utilized in the experiment. Participants and robot face each other. The teaching objects are located between the two interactants on a table. Upper left: *Looking around* behavior: no object is being presented to the robot. Upper right: *Reaching* behavior: triggered by a participant's presentation of an object with positive valence. Lower left, *Avoidance* behavior: triggered by a participant's presentation of an object with negative valence. Lower right: Modified *reaching* behavior, executed in prohibition experiment, where participants can push back the robot's arm to prevent it from touching a forbidden object.

participants. All other existing approaches on symbol grounding in developmental robotics do not make use of naïve participants, and appear to avoid unconstrained speech. There, the simultaneity constraint is typically enforced by either using trained participants, or system designers that are aware of the systems's technical limitations and the underlying assumptions. In either case the human interlocutors produce the right kind of words or utterances the right time in line with the system requirements. We refer to this type of interaction as "designer-robot interaction."

Another property of intent interpretations appears to be that they involve a certain asymmetry between the speakers: one party is conversationally leading or stronger, typically the mother or the teacher-participant, while the other party is conversationally weaker or inept such as the infant or the child-humanoid.

2.4 Overview of Prohibition Experiment

The prohibition experiment was an attempt to test Spitz' hypothesis in a human-robot interaction setup: Do linguistic prohibitions on part of the language tutors constitute likely candidates as functional origins of negation words? Instead of testing the hypothesis in isolation, we decided to test it alongside the hypothesis of the rejection experiment on intent interpretations as likely origins. The prohibition experiment is a genuine extension of the rejection experiment: both the robot's behavior as well as participants' task from the rejection experiment are taken as baseline, and are subsequently extended to elict linguistic prohibitions from participants.

3 MATERIALS AND METHODS

3.1 Study Design

As mentioned earlier, the *prohibition experiment* was designed in close alignment with the *rejection experiment* [14]. The robot's behavior developed for the rejection experiment formed the baseline behavior for the rejective scenario of the prohibition experiment (Figure 2). The cognitive and behavioral architecture employed was identical to the one employed in the rejection experiment with an extension that enabled the robot to cope with and react to being physically restrained. Every instantiation of the prohibition experiment consisted of five interactive sessions of approximately 5 minutes each. The multi-session format facilitated linguistic development over time and was required for the purpose of post-processing of participants' speech recordings.

Participants and robot were seated at opposite ends of a small table that contained several marked boxes and that constituted the to-be-taught objects (see Figure 1). Participants were told to teach the robot the names of these objects. The objects had randomly assigned valences for the robot that participants were initially not aware of. These would modulate the robot's motivation system and thereby trigger different behaviors (see next section).

Participants' speech was recorded via headsets and each session was video-recorded. Participants were not alone with the robot but either one or two more people were present. An operator was required to monitor the robot, and most often a helper was present to put boxes back on the table that had been dropped by the iCub robot Deechee. In few sessions the helper was absent such that the operator had to perform both tasks. As depicted in Figure 1, participant and robot were facing each other with a table separating the two. The teaching objects were 10-cm-long cardboard cubes that had black-and-white depictions of various shapes glued to each side of each box. The shapes were a square, a triangle, a star, a heart, and a crescent moon and all sides of a box showed identical shapes. Participants first read the instructions and signed the consent form. Afterwards they took their seat opposite the iCub and were subsequently asked to count down "three, two, one, start." So "start" acted as start marker for the session. After approximately five minutes the operator would give participants a signal. The operator, upon hearing "start," would press a button. This led to a time stamp being broadcasted through the architecture, which was recorded by the robot's body memory.

3.2 Robot Behavior

Both experiments utilized a motivation system that modulated the robot's affective expressions. These expressions were the facial expressions *smiling*, *frowning*, *neutral*, and matching body behaviors. Apart from modulating the robot's body behavior, its affective state was also fed into the symbol grounding system. As was the case in the *rejection experiment*, the robot's affective states were triggered by above mentioned object valences towards objects that were presented to it. The robot would attempt to grasp the object, in the case of a positive valence, or avoid it in the case of a negative valence. In addition, and differing from the *rejection experiment*, the robot's motivational state would also turn negative if it experienced external restriction of its arm movement. In absence of the aforementioned motivational triggers, the robot would display a baseline behavior during which it would alternate its gaze between the present objects and the participant's face.

If participants presented an object with a positive valence to the robot, then it would smile and hold out its hand towards them. It presented its palm, which signaled to participants that they could give it an object if they chose to do so (*reaching* behavior).

			Prohibition Experiment	Rejection Experiment
			Group 1 (Test Group)	Group 2 (Control Group)
sessions 1-3			*Prohibition + Rejection*	*Rejection only*
	Robot		{L D D L N}	{L D D L N}
	Human		{A A P P A}	{A A A A A}
sessions 4+5			*Rejection only*	*Rejection only*
	Robot		{L D D L N}	{L D D L N}
	Human		{A A A A A}	{A A A A A}

Fig. 2. **Alignment of Prohibition and Rejection Experiment**: *Prohibition* and *Rejection* in the table refer to the prohibition and rejection scenarios, respectively. Note that the *prohibition experiment* is composed of both the *prohibition* and *rejection scenarios*, whereas the *rejection experiment* consists of the *rejection scenario* only. Brackets (*{..}*) indicate the permutations of the following values: the robot either "liked" (*L*), "disliked" (*D*), or "felt"' neutral (*N*) about an object. Only participants knew whether an object was allowed *A* or prohibited/forbidden (*P*) for the robot to touch. The mapping of *positive*/*negative*/*neutral* (+/-/0) valences to objects was permuted between sessions, such that each object was twice *liked*, twice *disliked*, and once *neutral* across the 5 sessions (see Table 1). The mappings were identical for every participant. The *allowed*/*forbidden* markers were permuted as well.

If a participant presented an object with negative valence to the robot, then it looked at the object briefly, started to frown, and turned its head away. The dynamics of the interaction can lead to several consecutive "turn away" movements, which some participants interpreted as a form of head shake.

When presented with an object with neutral valence, the iCub displayed a neutral facial expression. Additionally it looked in regular intervals at both the participant's face and the presented object without approaching it (*watching* behavior). Prior to participants selecting the first object, but also between object presentations, the robot displayed a neutral face and switched its focus between all objects and the participant's face (*looking around* behavior).

The major driver of the robot's behaviors is thus its motivation system. The motivation system, in turn, is modulated by the external object valences as well as physical restriction of its arm movement. The latter has an exclusively negative impact and there is no positive counterpart.

The duration of the robot's gaze on a particular target were variable, and the time constants controlling the robot's gaze behavior were set to ones listed in the supporting materials (SM, Table 3).

During the first 3 sessions, approximately half of the objects were additionally tagged with markers, and participants were instructed to prevent the robot from touching these—the *prohibition task*. The last 2 sessions of the prohibition experiment are identical to the *rejection setup* such that the robot's success in acquiring negation words could be compared between the two experiments.

3.2.1 Rejection Scenario. The robot's behavior in the rejection scenario is the general behavior described above, but without extensions for the prohibition task. To increase the likelihood of the production of *negative intent interpretations*, we permuted the object-valence mapping for each session (see Table 1). In this way it was impossible for participants to know which object the robot would dislike at the outset of a particular session. This usually meant that they would present every disliked object at least once.

Table 1. Object Valences Per Session for Both Prohibition and Rejection Experiment

	session 1	session 2	session 3	session 4	session 5
triangle	1	−1	1	−1	0
moon	0	1	−1	1	−1
square	−1	0	1	−1	1
heart	1	−1	0	1	−1
circle	−1	1	−1	0	1

3.2.2 Prohibition Scenario. Extending the rejective scenario the prohibition scenario includes the prohibition task to elicit prohibitive utterances from participants. To this purpose two or three of the five present objects were declared to be *forbidden*, and were marked with colored dots on the side facing the participants. In addition to the instructions from the rejection scenario participants were told that the marked objects were forbidden objects, and that Deechee was not allowed to touch them. To keep Deechee from touching these forbidden boxes, participants were instructed to physically restrain the robot, in case it tried to touch them. Before the first session, participants were shown how to push the robot's arm back, first, to show them the ideal contact point, such that the robot's hand would not be damaged, and second, to take away their potential fear from actually touching the robot. The ideal contact point is the wrist and forearm. In this scenario, force control was used as the control mode for both arms, which makes it possible for participants to manipulate the arms while the robot executes a movement [31]. The act of pushing the robot's arm is detected as physical resistance and registered by the perception system as *resistance event*. The occurrence of such a resistance event leads to the robot's motivation being set to negative: Deechee subsequently starts to frown. Furthermore the face-related gaze time is increased to give this emotional display a slightly higher intensity (see SM Table 3).

The assignment of the forbidden and allowed attributes was such that every combination of *liked*/*disliked* with *allowed*/*forbidden* would occur at least once within each session. This, together with the change of the valence-to-object mapping between subsequent sessions (Table 1) then led automatically to a permutation of the *allowed* and *disallowed* attribute-to-object mappings across sessions. In general, there were either two or three *forbidden* objects per session, and two or three *allowed* ones.

In terms of the role reversal mentioned by Spitz none was needed as, in our architecture, there is neither a concept of self nor of other. For the prohibition experiment the behavioral trigger mechanism was modified such that negative motivation caused by participants' restriction of the robot's arm movement would not trigger its avoidance behavior. In these cases the robot would nevertheless frown in accordance with its motivational state.

3.3 Speech Processing, Symbol Grounding, and Speech-related Robot Behavior

The processing of participants' speech, as well as the symbol grounding was performed in between experimental sessions. The recorded speech was transcribed and speech recording and transcript manually re-aligned as the automatic alignment was not precise enough for the subsequent prosody recognition. Both participants' speech and the recorded sensorimotor and motivation (*smm*) data were timestamped to allow for temporal alignment in the symbol grounding phase.

From every utterance the prosodically most salient word was extracted, and the co-occurring sensorimotor-motivational (*smm*) data was attached to the word. At this point the temporal simultaneity mentioned in Section 2 becomes crucial: Those parts of the *smm* data are considered *relevant*, which were recorded during the production of the respective utterance. From this point on, the salient word is grounded in the robot's embodied "experience." After all of the participant's utterances had been processed in this manner, the new set of so grounded words is added to the robot's embodied lexicon. Note that a separate lexicon is created for each participant such that the robot follows an independent acquisition trajectory for every participant. The fact that it starts with an empty lexicon for each participant means that no designer knowledge in terms of a set of preselected words is incorporated. Everything that the robot eventually said during the experiment originated from what the respective participant had uttered in earlier sessions.

In every session but the first the lexicon was loaded into a memory-based learning system [8]. Importantly, the robot is real-time deaf, i.e., no real-time speech detection is in place. Rather than answering to questions, certain trigger behaviors will make the robot query its embodied lexicon. Trigger behaviors are behaviors which are caused by object presentations: *grasping* for, *rejecting*, or simply *watching* a presented object. Whenever the robot is engaged in one of these trigger behaviors, it matches its *smm* state against those associated to words in its current embodied lexicon and retrieves the best match. While a trigger behavior is active this process is continuously performed, which means at about 30 Hertz. As it is both impossible and impractical to speak at such a high rate, and because we expect a certain level of noise within the *smm* data, a thresholding mechanism is employed. This mechanism both stabilizes its linguistic output with respect to noise as well as adjusts its speech frequency to a more plausible level. The thresholding mechanism maintains a score for each best match word: The score of the respective word is increased whereas all other word scores are decreased. Once the score of a particular word reaches a certain threshold, the word is sent to the speech synthesizer: The robot speaks. After it spoke, all scores are reset to 0. The retrieval process now starts anew but on a reduced lexicon: the just-synthesized word has been removed. The removed word is added back to the lexicon once another word has been synthesized but also if a change in the *smm* state occurs.

3.4 Recruiting and Distribution of Participants

We recruited 10 participants all of whom were native English speakers and which were gender-balanced. The majority of participants were students or university employees and were naïve with respect to the true purpose of the experiment. In other words, the participants were unaware that linguistic negation was the topic under investigation. They were remunerated with £20 once they completed all five sessions. The protocol was approved by the ethics committee of the University of Hertfordshire under protocol number 0809/88, and approval was extended under protocol number 1112/42.

3.5 Instructions to Participants

To increase the likelihood of participants employing a speech register akin to child-directed speech, we asked them to imagine the robot as a pre-linguistic child. We also stated that the robot would like certain objects and dislike others. We used identical instructions in both the rejection [14] as well as the present prohibition experiment both of which, in turn, were very similar to *Saunders' experiment* [38]. The majority of instructions in both negation experiments as well as *Saunders' experiment* [38] were identical: participants were told that they ought to teach the robot Deechee about the objects on the table. We tried to prime participants into adopting a style of speech akin to child-directed speech (*CDS*, [15, 27, 43]) by telling them to imagine Deechee to be a two-year-old. In addition to what participants were told in *Saunders' experiment*, we also decided to mention to them that Deechee had likes and dislikes with respect to the objects such that they would not be caught by surprise by Deechee's emotional displays. It is not clear whether the latter instruction was strictly necessary, and whether it had any impact in terms of the content and style of participants'

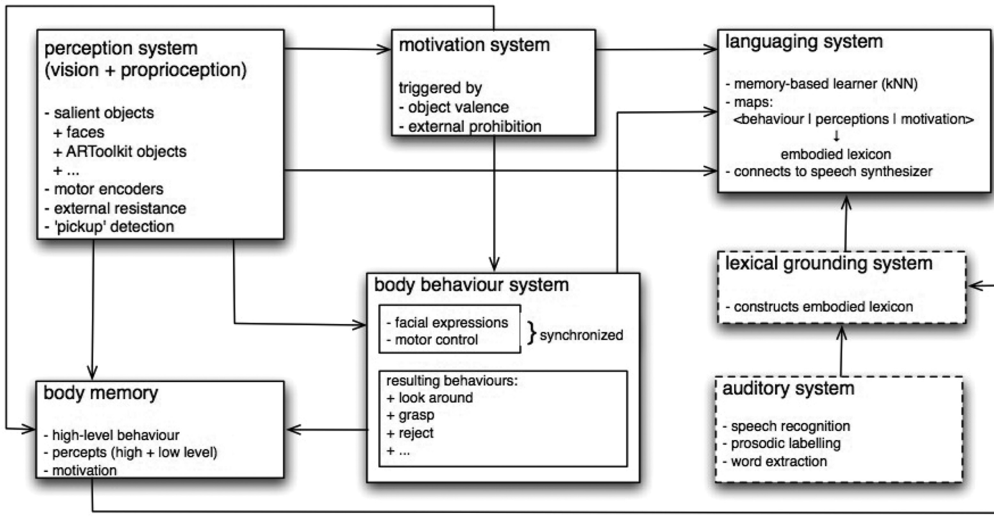


Fig. 3. **Functional overview of robotic architecture for language acquisition**. Solid lines indicate components that are active during experimental sessions ("online"), dotted lines indicate components that work offline.

speech. Only in the *prohibition experiment* participants were told that the marked objects on the table were forbidden for Deechee to touch. Participants were instructed to push the robot's arm away if it should approach these objects and they were practically shown how to do this.

3.6 Behavioral Architecture

The behavior architecture that generates both the humanoid's bodily and linguistic behaviour consists of the components depicted in Figure 3. We will sketch each component's purpose only very shortly as more elaborate descriptions have already been provided in References [12, 14].

The **perception system** gathers and processes percepts of all modalities. Visual processing was limited to face and object detection and based on the system developed by Rüsch et al. [36]. We also developed a detector for object-related *pick up* actions.

The **motivation system** is responsible for generating the affective-motivational state of the robot, which consists of a simple scalar value between −1 and 1: −1 corresponds to a negative, +1 a positive, and a small band around 0 a neutral state (cf. Varela et al. [52, Chapter 6]).

The **body behavior system** generates the humanoid robot's physical behavior, which also includes its facial expressions. The behavior is generated contingent upon the inputs from the perception and the motivation system. The behaviors are *Rejecting*, *Watching*, *Looking around*, *Reaching for object*, and *Idle*. Other subsystems are informed of changes in the bodily behavior by the broadcasting of unique behavior ids. Relevant time constants for certain parts of the behavior such as the eye gaze are listed in Table 3 of the supporting materials.

The **body memory** saves high-level and low-level perceptual data as well as behavior ids and the robot's motivational state to a file.

The **auditory system** encompasses speech recognition, word alignment, prosodic labeling, and word extraction, and are based on Saunders' system [40]. Utterance boundaries are set based on statistics of inter-word pause durations and word durations (see Saunders et al. [40] for details). Important for the later analysis is the notion of *prosodic saliency*. Note in this context that the first three aforementioned subsystems produce a sequence of utterances, where each utterance consists

Fig. 4. **Sensorimotor-motivational (*smm*) vector**. Solid lines mark those data dimensions that were used for symbol grounding and matching; *bid*: behavior id, *oid*: object id, *faceDet*: face detected, *moti*: motivation value, *resist*: resistance detected, *encX*: encoder #X.

of prosodically annotated words. Exactly one word is extracted per utterance, and that word is the prosodically most salient one. Prosodic salience is calculated as $f_0 * energy * d_w$, where f_0 is the maximum fundamental frequency, *energy* is the maximum energy, d_w is the word duration, and all of these components are normalized before said formula is applied.

The **lexical grounding system** performs the association of *smm* data with the salient words originating from the auditory system (see Figure 9). The so grounded words are subsequently added to the embodied lexicon.

The **languaging system** generates the robot's speech. It does so based on a process that matches the robot's current *smm* state (Figure 4) against the *smm* states associated to words in the embodied lexicon yielding a best-matching word in combination with a thresholding mechanism. At the core of the matching process is the k-nearest neighbor implementation Tilburg Memory-based Learner [50]. A new matching is performed whenever a new *smm*-vector is available, which is the case approximately every 30ms. The repetitive uttering of the same word is prevented by the use of a so called *differential lexicon*, which prevents the repeated production of the most recently uttered word. For details of both the thresholding mechanism and the *differential lexicon*, see Förster et al. [12, 14].

4 DATA ANALYSIS AND RESEARCH HYPOTHESES

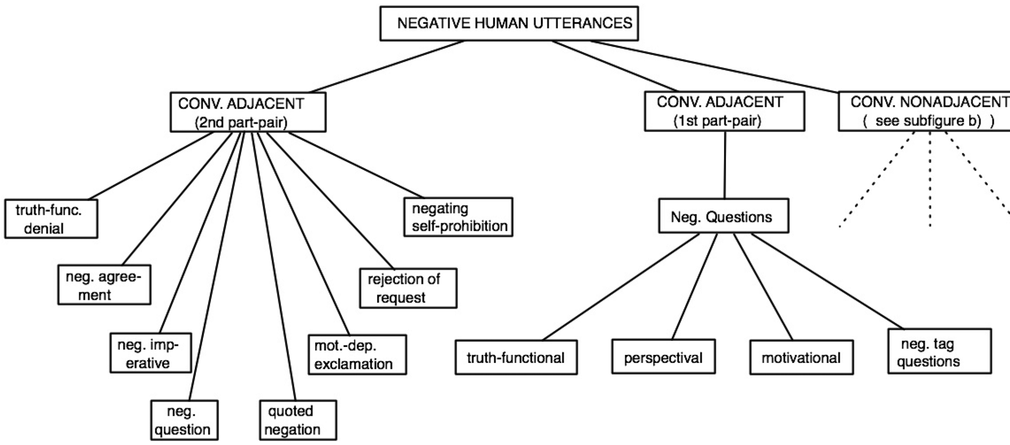
The independent variables in our experimental setup are the constitutive conditions of the two experiments: the presence of *rejective behavior and display* on part of the robot in the *rejection experiment*, and, additionally, the *prohibition task* and the accompanying body behavior of the robot in the *prohibition experiment*.

The data analysis, the results of which are reported in Section 5, is based on approximately 5 hours of participants' speech originating from 50 sessions—10 participants with 5 sessions each, gathered within the prohibition experiment (cf. SM Table 10). As was the case for the rejection experiment, analyses were performed on the word or corpus level, the utterance level, and the pragmatic level for negative words and utterances. Each level of analysis yielded dependent variables linked to one of the research hypothesis formulated at the end of this section. The dependent variables that will be used within the hypotheses are in the following marked as such ("***DV***"). The analysis will only be sketched as a more elaborate description has already been given in Reference [14].

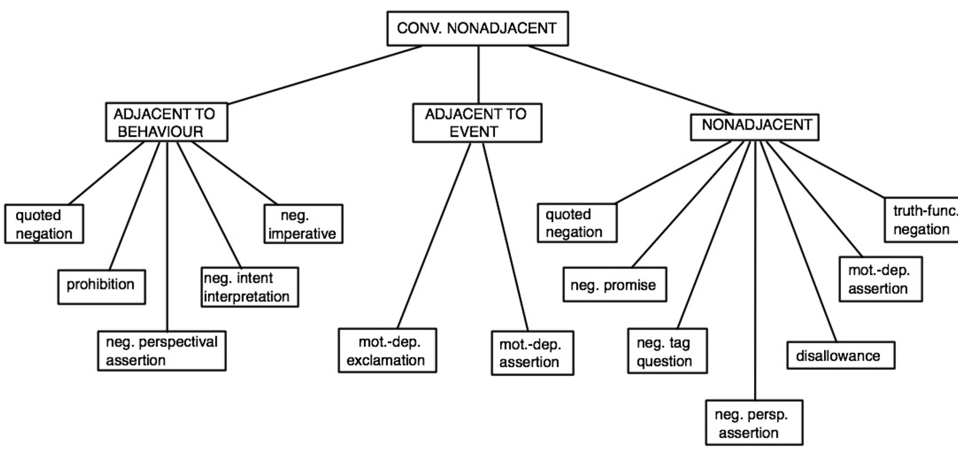
The following variables were measured on the level of utterances: speech frequency in utterances per minute (*u/min*), mean length of utterance (*MLU*), and the number of distinct words (cf. Figure 7 and SM Table 10).

More important for the purpose of understanding negation, are the counts of negative utterances per minute (*nu/min*, ***DV***), where negative utterances are utterances that contain at least one negative word. Whether a word is a negation word was determined manually by examining the complete list of distinct words compiled from all participants' speech transcripts.

On the *corpus-level* the prohibition corpus (*PC*) is a list of all words and their frequencies that occur in participants' speech transcripts taken from the prohibition experiment. The *PCS* is a subset of the *PC* containing only those words that were marked prosodically salient. Both corpora are



(a) Human Negation Types pt. 1



(b) Human Negation Types pt. 2

Fig. 5. **Taxonomy of negation types used by participants**. Conv.: conversationally, 1st part-pair, 2nd part-pair: parts of an adjacency pair such as question (1st part-pair)-answer (2nd part-pair).

presented together with the corresponding corpora from the *rejection* and *Saunders' experiment*: *RC*, *SC*, *RCS*, and *SCS* (cf. Figure 7(C)). On this level the most relevant number are the frequency ranks of negation words (***DV***), predominantly "no" and "don't," within the corpora of prosodically salient words. Only prosodically salient words enter the robots embodied lexicon. To be able to classify negative utterances of both participants and robot by their communicative function, their *pragmatic type* of sorts, we constructed two taxonomies, Figures 5 and 6. The construction process is described in Förster et al. [12, 14], but we emphasize that the resulting types can be regarded as types of speech acts in a loose sense, which were enriched by the notion of conversational adjacency. Conversational adjacency is not part of classical speech act theory [3, 41]. A short sketch of the most important negation types is given below, but for a detailed description of all types, we refer to the coding scheme [13]. Upon completion of the two taxonomies two coders classified the negative utterances by type, where the first coder classified all utterances, and the second coder classified a randomly selected subset comprising 20% of all utterances. This enabled

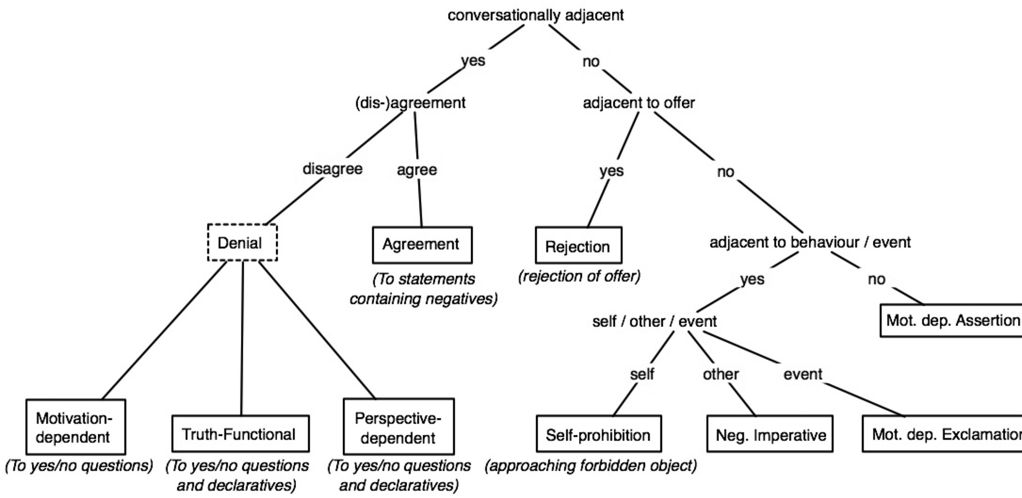


Fig. 6. **Taxonomy of robot negation types**. Types of negative utterances produced by the robot and as identified by external coders.

us to assess the taxonomies for internal consistency using Cohen's κ. Prior to coding the negative utterances for type, required for identifying intent interpretations, or prohibitions (***DV***), amongst other negation types, the coders coded the robots' utterances for felicity (***DV***). This means that they had to make a judgment whether they, by virtue of being fluent English speakers, perceive the negative robot utterance to be adequate or plausible in the respective conversational context. The internal consistency of the human taxonomy in terms of Cohen's κ was judged to be good ($\kappa = 0.74$), but the consistency of the robot taxonomy was judged to be only borderline acceptable, which triggered an automatic attempt to optimize it. Both the optimization attempt as well as our reasons for not adopting the recommended mergers suggested by the optimization are described in Förster et al. [12, 14]. Important for our current purposes is the fact that the κ values for the ratings of both the robot's felicity ($\kappa = 0.46$) and type ($\kappa = 0.41$) are at the very lower end of what is generally regarded acceptable. This has to be kept in mind when interpreting numbers that are based on these ratings. Importantly, however, there was no indication that any one of the two coders would have judged the robot's negative speech systematically more favorably as compared to the other (see SM Table 3).

Prior to describing the outcome of these attempts, we need to introduce the negation types mentioned within the present article. These include the ones most frequently produced during the experiments. A complete listing of all observed negation types can be found in the coding scheme [13]. In the following those types typically found in human participants' speech are qualified with "[H]." The types typically found in the robot's speech are marked "[R]." In the examples question marks indicate the intonational contour of a question, full stops the contour of an assertion.

Negative Intent Interpretations (NII [H]**)** are negative interpretations or ascriptions with regards to the addressee's motivational, emotional, or volitional state [33, p. 179].

Example utterances falling into this category are "No, you don't like fish" or a simple "No" if it is not produced as a genuine question.

Negative Motivational Questions (NMQ [H]**)** are very similar to *NII*s in that they refer to the addressee's perceived negative motivational state. The main difference between *NII*s and *NMQ*s is the fact that the latter are considered genuine questions, meaning, the speaker does expect the addressee to respond. Examples would be “Are you not feeling well today?” or “You don’t like apples?” in the context of being offered an apple, rather than the statement of a general preference.

Truth-functional Denials (TFD [H]**)** are used to deny a truth-functional assertion, with truth-functional assertions being assertions whose truth is independent of either speaker’s preferences or capabilities. Examples are “No, it’s not a hedgehog!” in the presence of an unknown animal and counter the suggestion of some other speaker or, again, a simple “No” in reply to some positive assertion.

Truth-functional Negations (TFN [H]**)** in our taxonomy are a catch-all category for all of those kinds of truth-functional negation that are not *truth-functional denials*, such as truth-functional suggestions or speculations, but also negative normative assertions such as, “In England you mustn’t drive on the right-hand side.”

Prohibitions (P [H]**)** are negative utterances whose function is to prevent the addressee from doing something. Considered in isolation, such utterances may not indicate that their function was prohibitive (cf. example 2). Taken out of context this utterance may be taken to be a truth-functional negation. However, in context, when looking at a video recording of the actual interaction, or when witnessing the latter “*in vivo*,” it becomes clear that this utterance is used as prohibition. In our experiment the prohibitive utterance can or cannot be accompanied by the participant physically restraining the robot’s arm movement.

Example 1: “No, you can’t touch that.”

Example 2: “No you’re not holding it, but you can look at them.”

Disallowance (D [H]**)** Disallowances are similar to *prohibitions* but, in contrast, capture those utterances that express general negative rules. In this sense disallowance utterances are more detached from the here and now of the interaction than *prohibitive* utterances. Whereas prohibitive utterances are always triggered by a current action on part of the robot, *disallowances* can or cannot accompany such an action.

Example: Speaker A takes something from the shelf and shows it to the robot saying “You can’t have this one.”

Negative agreements (A [H+R]**)** are negative confirmations in response to a negative statement such as a “no,” uttered by speaker 2 in response or addition to a “So you don’t like peanut butter, hmm?” by speaker 1.

Motivation-dependent Denials [R] are negative answers to *motivation-dependent questions* or *assertions*. Their content is dependent on the current emotional or volitional state of the addressee or her current preferences.

Example: “No” in response to “Do you want some ice cream?”

Rejections [R] are very close to *motivation-dependent denials*. The difference is that the latter is adjacent to an utterance whereas the former is adjacent or in reaction to non-linguistic offers. For example, “no” in response to someone holding out an apple as a offer would fall into this category.

Negative tag question (NTQ [H]**)** are negative clauses that are attached to the end of the utterance. Semantically and pragmatically they are probably the “least negative” of our negation types but they are easy to spot and appear to be highly frequent in British English. For example “don’t you” as in “You do like it, don’t you?” falls into this category. Another example would be “haven’t you” in “You have been to Cambridge, haven’t you?”

Once the speech recordings had been processed in this manner, we calculated statistics, where possible, such as the number of the various negation types per experiment. By linking the results of the pragmatic coding with the data that arose from the speech processing and symbol grounding, we could also start to answer questions about the frequency of certain negation words within the various negation types, and their prosodic saliency. In addition to the linguistic analyses a further analysis on the temporal relationships between participants' linguistic prohibition and their use of bodily measures to restrain the robot was performed (cf. SM Section A.4.1).

4.1 Research Hypotheses

Based on the introduced experimental variables, and given the research questions from Section 1.1.2, the research hypotheses are as follows. The research questions each hypothesis links up with are stated in brackets at the end.

- (H1) Human tutors, when confronted with a defiant humanoid robot, and in the presence of forbidden objects, produce more linguistic prohibitions as compared to the rejection scenario where such objects are absent. (~Q2)
- (H2) Assuming that *H1* holds, prohibitions contain at least as many negation words as do negative intent interpretations. (~Q3, Q4)
- (H3) Assuming that *H1* holds, the prosodic saliency of negation words within prohibitions is as high or higher than the one observed in *negative intent interpretations*. (~Q3, Q4)
- (H4) Assuming that *H1* holds, the felicity rate of the robot's negative utterances will be higher than what was observed in the rejection experiment. (~Q5)

5 RESULTS

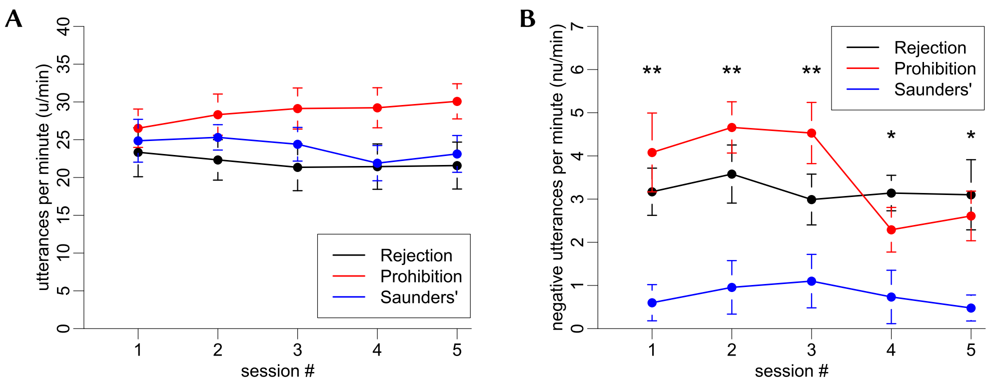
In the prohibition experiment, on average, every seventh to eighth utterance of participants contains a negation word, which constitutes an increase of 391% compared to the speech recorded in Saunders' scenario (cf. Figure 7). This compares to a rise of 332% in the rejection experiment. The frequent occurrence of negative utterances leads to a large increase of prosodically salient negation words, which subsequently enter the robot's active vocabulary. The increase is amplified by the relatively high prosodic saliency rate of "no" in both negation experiments. The prosodic saliency of negation words, chiefly "no," is high both in relation to Saunders' experiment as well as in relation to the average word saliency within the negation experiments (Figure 8(A)). As a consequence 'no' rises to the fourth rank in the prohibition corpus of salient words *PCS*, and even to second rank in the corresponding *RCS* (Figure 7).

Hypothesis 1. Analysing these negative utterances with respect to their communicative function reveals that, within the prohibition experiment, linguistic *prohibitions*, not present within the rejection experiment, occupy the top-rank, making up 21% of all negative utterances. Our results hence corroborate *hypothesis 1*: our participants did indeed produce a multitude of linguistic prohibitions in response to the presence of the prohibition task.

The high rate of prohibitions relative to all produced negation types is remarkable as linguistic *prohibitions* were only produced when the prohibition task was given, i.e., during the first three sessions, whereas utterances of the other types were produced in all of the five sessions. *Prohibitions* are followed by *negative intent interpretations* (18%), *negative motivational questions* (18%), and *truth-functional denials* (11%) (cf. Figure 8 and SM Table 4), with prosodic saliency rates of 60.5%, 38.2%, 41.8%, and 31.7%, respectively.

This means the three motivation-dependent types provide the majority of negation words for the robot's lexicon due to two factors: First, these types of utterances are dominant in terms of the absolute numbers of productions, and, second, the negation words that are part of these utterances have higher rates of prosodic saliency than the truth-functional types.

In comparison, within the rejection experiment, at 31% the most frequent negation type, are *negative intent interpretations* (*NII*), followed by *negative motivational questions* (*NMQ*, 30%). Both of these types have a direct link to the robot's display of affect. *Truth-functional denials* (*TFD*) rank third (22%) in the *RC* but have a lower saliency rate (29%), relative to the two motivational



C

	All words						Salient words only					
	Rej. Corpus		Pro. Corpus		S. Corpus		Rej. Corpus		Pro. Corpus		S. Corpus	
rank	word	%	word	%	word	%	word	%	word	%	word	%
1	you	7.13	you	6.18	a	8.71	square	4.97	square	4.4	blue	6.91
2	the	5.63	the	5.5	this	4.55	no	4.64	triangle	4.06	red	5.54
3	like	3.31	a	3.74	blue	4.31	triangle	3.95	circle	3.83	circle	5.15
4	a	2.72	this	3.71	is	4	heart	3.8	no	3.66	heart	4.75
5	this	2.7	one	2.8	and	3.9	moon	3.53	one	3.36	green	4.36
6	no	2.39	is	2.45	red	3.75	circle ↑	3.53	heart	3.34	arrow	3.56
7	one	2.27	like	2.05	green	3.55	like	3.2	this	3.09	cross	3.48
8	square	1.93	to	1.83	the	3.29	circles	2.69	moon	2.8	side ↑	3.48
9	do ↑	1.93	no	1.79	that's	2.54	squares	2.42	ok	2.3	box	2.82
10	to	1.78	it's	1.66	you	2.41	it	2.36	shape	2.14	shape	2.42
11	it ↑	1.78	heart	1.6	it's	2	yes	2.28	yes	2.09	and	2.11
12	that	1.73	square	1.51	heart	1.99	one	2.13	crescent	2	moon ↑	2.11
13	moon	1.62	triangle	1.47	circle	1.85	right	1.88	like	1.8	square	2.07
14	heart	1.6	that	1.42	arrow	1.84	this	1.82	circles	1.74	this	2.02
15	is	1.47	it ↑	1.42	side	1.81	ok	1.73	again	1.51	star	1.85
16	triangle	1.45	do	1.4	cross	1.49	again	1.69	good	1.44	ok	1.76
17	circle	1.32	moon	1.38	here	1.39	ok	1.57	it	1.36	is ↑	1.76
18	it's	1.22	circle	1.29	we	1.38	good	1.55	very	1.25	small	1.54
19	don't	1.15	that's	1.28	on	1.37	Deechee	1.46	ok	1.18	right ↑	1.54
+1	see	1.12	shape	1.2	no (*50*)	0.35	Deechee[2]	1.3	Deechee	1.12	no (*32*)	0.44

Fig. 7. Impact of motivated behavior on linguistic production of participants. (A) The overall production rates between prohibition (upper red), rejection (lower black), and Saunders' experiment (middle blue) differ only marginally (mean ± SEM). (B) The production rate of negative utterances only, however, is significantly higher in the prohibition (upper red) and rejection (middle black) as compared to Saunders' experiment. See *Supporting Information* for details. $*p < 0.05$, $**p < 0.01$ (C) The large supply of negative utterances has consequences on the corpus level: *No* is amongst the 10 most frequent words in the rejection and the prohibition corpus, whereas it is located on rank 50 in Saunders' corpus. In the corpora of prosodically salient words *no* ranks even higher. This is due to the high saliency of the word and it subsequently enters the robot's vocabulary frequently. (Arrows (↑) indicate equality of ranks between the stated entry and the next entry above. Negation words are marked through gray background. The "+1" row contains the 20th most-frequent words unless a different rank is specified in brackets.)

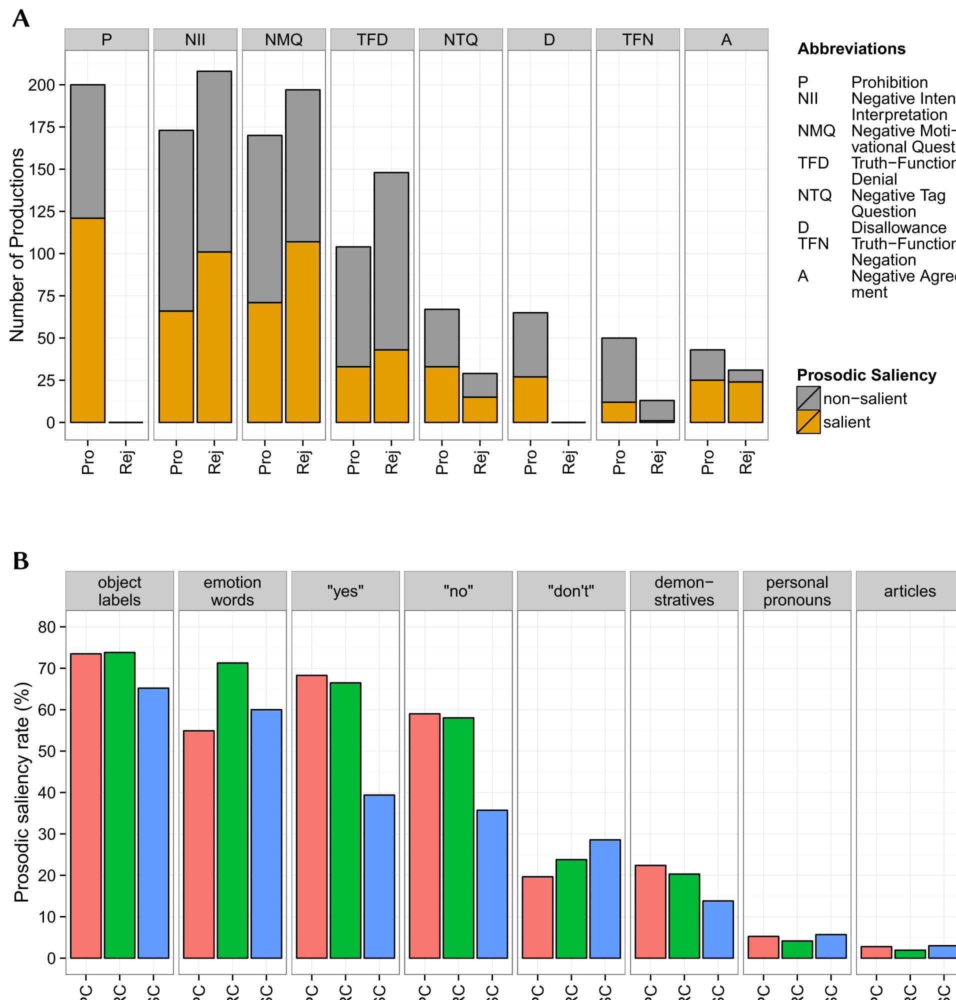


Fig. 8. (*A*) Frequency of human utterances classified as being of the stated negation types (pragmatic level) and percentage of utterances falling under the respective type with salient negation word (only types with >5% of total number of negative utterances, *Pro*: Prohibition Experiment, *Rej*: Rejection Experiment). (*B*) Prosodic saliency rates of selected words and word groups. "No" has a considerably higher salience rate in the two negation experiments as compared to Saunders' experiment. *PC*: Prohibition Corpus, *RC*: Rejection Corpus, *SC*: Saunders et al.'s Corpus.

types (*NII*: 54%, *NMQ*: 49%). From this we can conclude that in the *rejection experiment* the vast majority of negation words in the robot's active vocabulary originate from utterances of the two motivation-dependent types, *NII*s and *NMQ*s.

When comparing the two negation experiments, it becomes clear that within the prohibition experiment *negative intent interpretations* and *negative motivational questions* were produced less frequently than was the case in the rejection experiment. The probable cause for this is the fact that, in both experiments, participants had overall the same amount of time, yet participants in the prohibition experiment spent part of their time with attempts to prohibit the robot leaving them less time to engage in *NII*s and *NMQ*s.

Table 2. Statistical Comparison of Felicity Rates of iCub's Negative Utterances

Criterion	*Experiment*	*Felicity rate (%)* mean (std)	*Test result*
crit. 1	Rejection	74.40 (23.80)	$t(14) = 3.09, p = 0.008$
	Prohibition	32.57 (30.31)	
crit. 2	Rejection	64.16 (19.80)	$t(10) = 3.17, p = 0.013$
	Prohibition	28.02 (17.08)	

Given are the mean and standard deviation for the felicity rates of the robot's production of negation during sessions 4 and 5. Two criteria: **Crit. 1**: Data basis is felicity values of all but those where the robot did not produce any negative utterances (sessions with *P01*, *P06*, *P10*, and *P18*). **Crit. 2**: Data basis as in crit. 1 plus additional exclusion of those felicity rates based on less than 10 utterances, that is, additional exclusion of sessions with *P04*, *P09*, *P13*, *P19*, and *P22*. See Table 15 for a more detailed breakdown of the robot's success rates itemized by participant and negation type.

Hypothesis 3. This hypothesis is confirmed by the analysis of the experimental data. When considering the saliency rates of negation words within utterances of the aforementioned types produced within the prohibition experiment it becomes clear that linguistic *prohibitions* constitute a formidable source of negation words: within this type, negation words reach the overall highest saliency rate (61%), followed by *negative motivational questions* (41%), *negative intent interpretations* (38%), and *truth-functional denials* (32%) (see SM Table 5 for the complete listing). The combination of high production rate and high saliency rate renders them the top contributors of negation words to the robot's active vocabulary within this experiment.

Hypothesis 2. This hypothesis is being confirmed by the data analysis. Every single prohibitive utterance contained at least one negation word whereas intent interpretations and motivational questions were sometimes performed in a non-negative way. Some participants for example used non-negative emotion words in response to the robot's negative affective display such as 'sad' as in "why are you sad?" where others more commonly used the negative "you don't like it?" Thus, from a merely lexical perspective, *prohibitions* appear to be more reliable sources of negation words than *intent interpretations* and *motivational questions*.

In *Saunders' experiment* in comparison "no," is ranked 50th in the *SC* and 32nd in the *SCS*. There it accounts for less than 0.5% of words in both corpora. Hence, both the affective or motivational displays of the robot as well as the prohibition task lead to a considerably higher rate of negative utterances when compared to the setup used by Saunders et al., who used a near-identical setup but without affective displays and without prohibition task.

Hypothesis 4. This hypothesis has *not* been confirmed. To our surprise the robot's learning success with respect to negation was judged to be considerably lower in the prohibition (~28–33%) as compared to the rejection experiment (~65–74%, see Table 2).

This result triggered an additional analysis where we aligned the signal representing external pressure on the robot's right arm with both the robot's affective state as well as the timing of negative utterances of the four most frequent negation types, the details of which are provided in the supporting materials (Section A.4.1) This analysis showed that participants from the prohibition experiment 46% of the time did not physically restrain the robot's arm when uttering *prohibitions* as instructed, and in 12% of all cases uttered *prohibitions* before applying restraint (cf. SM Table 16). In both cases grounded words enter the robot's lexicon where the negation word is likely to be associated with positive affect (see SM Table 17). This may then lead to the inappropriate usage of the word. In only 18% of cases did our participants utter *prohibitions* while restraining the robots arm, that is, while it was in a negative affective state. When performing a similar analysis for *negative intent interpretations*, we observed that in approximately two thirds of cases (rejection experiment: 66%, prohibition experiment: 63%) the robot is in a negative motivational state as opposed to a positive one (rejection experiment: 9%, prohibition experiment: 6%) such that there is a high likelihood of it associating the negative word with negative affect. For *negative motivational questions* the results look similar within the prohibition experiment (58% performed while in a negative state, and only 9% while in a positive state), while in the rejection experiment the number of performances while in a negative state (40%) is nearly identical to the number of performances while in a neutral state (41%). Performances of this type while the robot is in a positive state are also not very frequent within this experiment (19%). Thus, albeit lexically not being equally reliable sources of negation words as *prohibitions*, *negative intent interpretations* appear to be better sources for the establishment of an association of the negative word with negative affect if word learning is mainly modeled as a process of establishing associations between sensorimotor-affective 'concepts' and linguistic items.

6 DISCUSSION

6.1 Research Questions Answered

6.1.1 Q1: Are Affect or Motivation Necessary Requirements in the Acquisition of Negation? Under the assumption that the constructive methodology that we have chosen is a valid means to explore genuinely human phenomena such as language, our experiments corroborate the assertion that "having" affect or motivation is a crucial ingredient when it comes to acquiring negation words. In both negation experiments the presence of affective or motivated behavior on part of the robot caused significant changes in human tutors' speech behavior. Where they had produced close to no negation words in a nearly identical tutoring setup in Saunders' experiments, their production was abundant in both negation experiments. In addition to the strong rise in the frequency of negation words, these words were also prosodically salient to a level that is matched only by object labels and emotion-related words. This makes them likely to be picked out from the speech stream by a budding language learner.

6.1.2 Q2: Do Human Participants Produce Prohibitions in a HRI-setup Akin to what Can be Observed in Parent-child Dyads? If the setup is similar to the one we have used, then the answer is a simple "yes." Participants' speech behavior was in this respect what one would intuitively expect in the given situation if the dyad would have been parent-child: it is abundant with prohibitions. As there is no quantitative data on the parental use of prohibitions in parent-child conversation, more detailed comparisons cannot be made at this point.

6.1.3 Q3: Are Prohibitions Good Sources for Negation Words? Prohibitions are excellent sources for negation words, for two reasons: Every single prohibition that we encountered contained a negation word, and in a large percentage of them a negation word was the prosodically salient word of that utterance.

6.1.4 Q4: How do Prohibitions Compare to Negative Intent Interpretations? While negative intent interpretations are already very good sources for negation words (cf. Förster et al. [14]), from a lexical and prosodic perspective, prohibitions are even better sources for the reasons stated in our answer to question 3.

6.1.5 Q5: Did the Robot's Felicity Rate Improve Due to Exposure to the Tutor's Prohibition? No, it did not. Compared to the rejection experiment the adequacy of the robot's use of negation worsened considerably. The main reason for this surprising lack of improvement appears to be that the simultaneity constraint mentioned in Section 2.3 had been frequently violated. In other words: the robot was not in a negative affective state at the time when prohibitions were produced by the participants. Whether or not this is equivalent to what would be observed with parent-child dyads is unknown due to the lack of sufficiently detailed comparative data in the child language literature. More generally, there is no quantitative psycholinguistic data that we could use to compare the robot's felicity rate to those of young children when it comes to the use of negation words. Documented in the literature are semantic mistakes during certain developmental stages such as the over-generalisation of nouns [16] and grammatical constructions [5, 6]. Unfortunately there is no such data for "pragmatic accuracy" in terms of the (non-)successful use of negation words such as "no."

Despite the lack of comparative data, we hypothesize that our behavioral setup and learning architecture were most likely somewhat too simplistic to handle the complexities of multimodal prohibitive behavior. Intuitively a 30% success rate appears to be very low. Here are the reasons why we think that our setup was too simplistic to "make proper use" of tutors' prohibitions.

6.1.6 Why did the Robot Fail to Learn? Our participants, despite having been instructed on how to physically prevent the robot from touching an object, often chose not to do so. Instead, we observed that participants frequently held forbidden objects out of the robot's reach instead of limiting its arm movement. As the robot does not perform any type of goal-evaluation with respect to having reached or touched the "desired" object it does not become frustrated in these cases. While we are unaware of any existing studies on children's acquisition of negation that would provide us with the level of interaction detail that we would require (cf. SM Section A.4.1), anecdotal evidence from our own children can give us some hints.

Having demonstrated to our participants how to physically keep the robot from touching a forbidden object, we somewhat naïvely expected that they would just do that whenever the robot engaged in such forbidden behavior. However, they did not. One explanation would be that our participants' were simply reluctant to touch the robot most of the time. In this case there would be a mismatch between the synthetic model, the human-robot dyad, and its natural counterpart, the parent-child dyad. A less innocent explanation in terms of the consequences for our behavioral design would be that people generally do not excert corporal restraint easily but rather use other means of keeping children from accessing what should not be accessed. One way of achieving this is by simply putting the target object out of the toddler's or robot's reach, which is what we have observed with some participants. Based on the aforementioned anecdotal evidence such moves are likely to frustrate the toddler. However, putting object's out of our robot's reach did not frustrate it due to relatively fundamental reasons: Hebbian-type algorithms do not make reference to goal-states. In our architecture, we modeled world learning to be by and large associative or Hebbian. This decision was not made on the basis on principle, but rather by the application of Occam's razor: Given the lack of evidence to the contrary, we chose one of the simplest types of learning algorithms. Yu et al. [54] provide a more elaborate discussion on associative learning in word acquisition.

One consequence of not having explicit goals on the level of the learning algorithm is that frustration, that is, a state caused by not reaching a goal, or not reaching it within a certain time span, is hard to model. In our architecture, frustration is not explictly modeled. Instead external physical restraint will directly trigger the robot's negative affect, without the "intermediary" notion of frustration.

6.1.7 Anecdotal Evidence on the Natural "Structure" of Prohibition. The lack of detail in the developmental literature with respect to the precise temporal unfolding of prohibition led us to search the video platform YouTube for amateur videos depicting parents that engage in prohibition.

Albeit of a somewhat anecdotal character some videos show situations where parents clearly prohibit their children by the mere use of speech and cause the child's frustration as a consequence of these "touch-less" acts of prohibition ("single whammy prohibition") [45]. In other videos, however, prohibiting utterances are swiftly followed by a combination of corporal restraint and linguistic prohibition akin to the behavior we expected our participants to engage in ("double whammy prohibition") [9]. Due to the anecdotal character of this evidence it is impossible to tell which of these two variants of prohibition is more typical, whether one is the developmental precursor of the other, or whether it is a matter of the severity of the violation rather than a matter of the developmental stage. It is not hard to imagine that a child, having been exposed to several instances of "double whammy" prohibition, would learn that "resistance is futile" and that, as a consequence of this learning process, "single-whammy" prohibition suffices to stop the child from engaging in the prohibited behavior (*variant A*).

However, it seems equally plausible that prohibitive utterances may carry distinctive prosodic features that let the child infer the caretaker's negative emotional or volitional stance. We may assume that a child in the relevant age range can perform simple inferences within a lay theory of emotions or volition [29, 30]. If we then further assume that the child experiences the caretaker's relative interactional and interventional power on a daily basis, then the child should be able to infer that resistance is futile without a need for prior exposure to "double whammy" prohibition (*variant B*). The lack of more than anecdotal evidence prevents us from excluding one or both of these possibilities.

6.1.8 Reinforcement Learning as Alternative? It may be of interest to observe that both *variants A* and *B* require a more powerful class of learning algorithms than the simple Hebbian-type one employed by us. Both of these variants assume that the agent learns about the efficacy of its actions with respect to some goal. In machine learning, this would be typically modeled with some type of reinforcement learning, which is arguably a more powerful class of learning algorithms than simple Hebbian-type associative learning. The replacement or supplement of our memory-based learner with some type of reinforcement learning as core learning mechanism and the explicit modeling of goals would therefore most certainly increase the felicity rate when using negation words. In this case the robot's frustration could be triggered whenever it cannot reach an object within a certain time frame and when this inability is caused by another agent. This would then lead to a rise in the number of grounded negation words with negative motivation value in the data set of the memory-based learner.

Under the assumption that prohibitive utterances are indeed the main source of children's early negation words, the only potential rescue for a purely associationistic account we can conceive of hinges on the notion of emotional contagion [19] (*variant C*). It shares with *variant A* the idea that acoustic or prosodic qualities of prohibitive utterances, potentially in conjunction with corresponding facial expressions, may carry an emotional charge. As we have seen from our data, prohibitions are typically prosodically salient. Assuming that their acoustic properties may have the potential to negatively impact the affective state of the recipient, corporal restraint might not be necessary to "turn the infant's mood sour." Corporal restraint may indeed only be used by the care-giver as the very last resort. If such a mechanism of acoustic affective contagion could be implemented within our learning architecture, then we would arrive at a point where the non-codified aspects of utterances would contribute to the modulation of an interlocutor's affective state, which in turn would form part of the basis for grounding the codified units of the same, or adjacent utterances.

The difference between this account of emotional contagion (*variant C*) and *variant A* above is that the former does not require any reasoning process operating on actions and goals. It ascribes to the parent the power to impact the child's emotional state more or less directly by producing utterances with a certain emotional charge. In the account under *variant A* the parents power to affect the child's motivational state is more indirect: The utterance's emotional charge, rather than impacting the child directly, is taken into account by the child's goal-oriented reasoning process. To be efficacious for grounding negative symbols this process must be social: the reasoning must not only take into account the child's own abilities and goals but also the abilities and goals of the caretaker, and evaluate whether a caretaker's intervention is likely if a certain action is chosen and given the emotional payload of the previously received utterances and other communicative emotional signals.

As can be seen from these considerations the degrees of freedom for potential modifications of the learning architecture are many. Only studies with temporal high-resolution and multimodal descriptions of parental prohibitive behavior have the potential to create the required comparative data set. Once available, such data set could then not only provide us with better means to evaluate our results but also reduce the degrees of freedom for future modifications of our learning architecture as indicated above.

In the context of the rejection experiment [14], we determined that a lack of proper timing, that is uttering a "no" after a pause longer than the important 1 second threshold [22], caused confusion in the coders when trying to determine the meaning of the word, and will presumably cause similar problems with the interactor. Similarly the choice between the hypothesized learning mechanisms in the context of prohibitive utterances could be informed by detailed observations of timing. Assuming that emotional appraisal processes are faster than inferential processes, the temporal order between prohibitive utterances and the child's overt emotional displays could provide clues as to which of the two types of processes is at play.

7 CONCLUSIONS

Our architecture is the first to extend symbol grounding beyond the realm of sensorimotor-data to encompass affect and motivation, which is in line with recent psychological studies [24]. We have demonstrated the capacity to acquire generally felicitous non-referential linguistic behavior such as negation in a developmental scenario with a humanoid robot developing an embodied lexicon based on its sensorimotor motivational experience in interaction with naïve human participants. Based on our results we cannot exclude any of the two hypotheses on the origin of negation, due to a lack of sufficiently detailed data on the precise dynamics how prohibition in mother-child dyads is enacted. We did show however, that at least from a lexical perspective prohibitive utterances are formidable sources of negation words.

If more detailed data on the dynamics of prohibition were to become available, then the results of the presented research can give strong indications with respect to the underlying acquisition algorithm: If the majority of prohibitive utterances are uttered at a time when the child is already frustrated, then Hebbian-style methods may suffice to ground negative words in negative affect. Yet our analysis on the temporal alignment between bodily and linguistic behavior hints towards principal limitations of Hebbian-style learning. If prohibitive utterances typically precede or may even cause a child's frustration, then Hebbian-style methods are unlikely to be efficacious for affective grounding, because they require a certain amount of synchronicity between negative word and negative affect. In this case, a more powerful type of learning algorithm would be required. Reinforcement learning, potentially coupled or amplified by some form of social reward signals would be a likely candidate class of learning algorithms. This view appears to be supported by recent work in neuroscience, which posits a central role of reinforcement learning in biological decision making [28], if we are willing to assume that language learning may recruit more general learning mechanisms.

ACKNOWLEDGMENT

The authors thank research assistant Kinga Grof for helping with the manual transcription and re-alignment of recorded speech.

REFERENCES

[1] David W. Aha (Ed.). 1997. *Lazy Learning*. Springer-Science+Business Media, Dordrecht.
[2] Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. 2009. Cognitive developmental robotics: A survey. *IEEE Trans. Auton. Mental Dev.* 1, 1 (2009), 12–34.
[3] John Langshaw Austin. 1975. *How to Do Things with Words.* Harvard University Press, Cambridge, MA.
[4] Lawrence W. Barsalou. 1999. Perceptual symbol systems. *Behav. Brain Sci.* 22, 4 (1999), 577–660.
[5] Melissa Bowerman. 1988. The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In *Explaining Language Universals*, J. Hawkins (Ed.). Basil Blackwell, Oxford, 73–101.
[6] Patricia J. Brooks, Michael Tomasello, Kelly Dodson, and Lawrence B. Lewis. 1999. Young children's overgeneralizations with fixed transitivity verbs. *Child Dev.* 70, 6 (1999), 1325–1337.
[7] Angelo Cangelosi. 2010. Grounding language in action and perception: From cognitive agents to humanoid robots. *Phys. Life Rev.* 7, 2 (2010), 139–151.
[8] Walter Daelemans and Antal van den Bosch. 2005. *Memory-based Language Processing*. Cambridge University Press, Cambridge, MA.
[9] Misti DeWeerd. 2011. Daddy says no. Retrieved from https://youtu.be/6b8dafQ5OQw.
[10] P. F. Dominey and J. D. Boucher. 2005. Learning to talk about events from narrated video in a construction grammar framework. *Artific. Intell.* 167 (2005), 31–61.
[11] Larry Fenson, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal, and Stephen J. Pethick. 1994. Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1 (1994), i–185.
[12] Frank Förster. 2013. *Robots that Say "No": Acquisition of Linguistic Behaviour in Interaction Games with Humans.* Ph.D. Dissertation, University of Hertfordshire.
[13] Frank Förster. 2018. *Coding Scheme for Negative Utterances.* Technical Report. Adaptive Systems Research Group, University of Hertfordshire. https://doi.org/10.18745/pb.20312
[14] Frank Förster, Joe Saunders, and Chrystopher L. Nehaniv. 2018. Robots that say "No": Affective symbol grounding and the case of intent interpretations. *IEEE Trans. Cogn. Dev. Syst.* 10, 3 (2018), 530–544. DOI: https://doi.org/10.1109/TCDS.2017.2752366
[15] Clare Gallaway and Brian J. Richards. 1994. *Input and Interaction in Language Acquisition.* Cambridge University Press, London.
[16] Susan A. Gelman, William Croft, Panfang Fu, Timothy Clausner, and Gail Gottfried. 1998. Why is a pomegranate an apple? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of apple and dog. *J. Child Lang.* 25, 2 (1998), 267–291.
[17] Alison Gopnik. 1988. Three types of early word: The emergence of social words, names and cognitive-relational words in the one-word stage and their relation to cognitive development. *First Lang.* 8, 22 (1988), 49–70.
[18] S. Harnad. 1990. The symbol grounding problem. *Physica D* 42 (1990), 335–346.
[19] Elaine Hatfield, John T. Cacioppo, and Richard L. Rapson. 1993. Emotional contagion. *Curr. Direct. Psychol. Sci.* 2, 3 (1993), 96–99.
[20] George Hollich, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2000. I. what does it take to learn a word? *Monogr. Soc. Res. Child Dev.* 65, 3 (2000), 1–16.
[21] Ian Hutchby and Robin Wooffitt. 1999. *Conversation Analysis: Principles, Practices, and Applications.* Blackwell, Malden, MA.
[22] Gail Jefferson. 1989. Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. In *Conversation: An Interdisciplinary Perspective*, Derek Roger and Peter Bull (Eds.). Multilingual Matters, Clevedon, Chapter 8, 166–196.
[23] Casey Kennington and Sarah Plane. 2017. Symbol, Conversational, and Societal Grounding with a Toy Robot. arXiv:1709.10486.
[24] Stavroula-Thaleia Kousta, Gabriella Vigliocco, David P. Vinson, Mark Andrews, and Elena Del Campo. 2011. The representation of abstract words: Why emotion matters. *J. Exper. Psychol.* 140, 1 (2011), 14–34.
[25] Caroline Lyon, Chrystopher L. Nehaniv, Joe Saunders, Tony Belpaeme, Ambra Bisio, Kerstin Fischer, Frank Förster, Hagen Lehmann, Giorgio Metta, Vishwanathan Mohan, et al. 2016. Embodied language learning and cognitive bootstrapping: Methods and design principles. *Int. J. Adv. Robot. Syst.* 13, 3 (2016), 105.
[26] Chrystopher L. Nehaniv, Frank Förster, Joe Saunders, Frank Broz, Elena Antonova, Hatice Köse, Caroline Lyon, Hagen Lehmann, Yo Sato, and Kerstin Dautenhahn. 2013. Interaction and experience in enactive intelligence and humanoid robotics. In *Proceedings of the IEEE Symposium on Artificial Life (ALIFE'13).* IEEE, 148–155.

[27] Elissa L. Newport. 1977. Motherese: The speech of mothers to young children. In *Cognitive Theory Vol. 2*, John N. Castellan, David B. Pisoni, and George R. Potts (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 177–217.
[28] Yael Niv. 2009. Reinforcement learning in the brain. *J. Math. Psychol.* 53, 3 (2009), 139–154.
[29] Desmond C. Ong, Jamil Zaki, and Noah D. Goodman. 2015. Affective cognition: Exploring lay theories of emotion. *Cognition* 143 (2015), 141–162.
[30] Desmond C. Ong, Jamil Zaki, and N. D. Goodman. 2019. Computational models of emotion inference in Theory of Mind: A review and roadmap. *Topics Cogn. Sci.* 11, 2 (2019), 338–357.
[31] Ugo Pattacini. 2010. *Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub*. Ph.D. Dissertation, RBCS, Istituto Italiano di Tecnologia, Genoa, Italy.
[32] Roy D. Pea. 1978. *The Development of Negation in Early Child Language*. Ph.D. Dissertation, University of Oxford.
[33] Roy D. Pea. 1980. The development of negation in early child language. In *The Social Foundations of Language & Thought: Essays in Honor of Jerome S. Bruner*, D. R. Olson (Ed.). W. W. Norton, New York, 156–186.
[34] D. Roy. 2005. Semiotic schemas: A framework for grounding language in action and perception. *Artific. Intell.* 167, 1–2 (2005), 170–205.
[35] Deb K. Roy and Alex P. Pentland. 2002. Learning words from sights and sounds: A computational model. *Cogn. Sci.* 26, 1 (2002), 113–146.
[36] Jonas Rüsch, Manuel Lopes, Alexandre Bernardino, Jonas Hörnstein, Jose Santos-Victor, and Rolf Pfeifer. 2008. Multimodal saliency-based bottom-up attention—A framework for the humanoid robot icub. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'08)*. IEEE, 962–967.
[37] Joanna Ryan. 1974. Early language development: Toward a communicational analysis. In *The Integration of a Child into a Social World*, M. P. M. Richards (Ed.). Cambridge University Press, London.
[38] Joe Saunders, Hagen Lehmann, Frank Förster, and Chrystopher L. Nehaniv. 2012. Robot acquisition of lexical meaning - moving towards the two-word stage. In *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob'12)*. IEEE, 1–7.
[39] Joe Saunders, Hagen Lehmann, Yo Sato, and Chrystopher L. Nehaniv. 2011. Towards using prosody to scaffold lexical meaning in robots. In *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob'11)*.
[40] Joe Saunders, Chrystopher L. Nehaniv, and Caroline Lyon. 2011. The acquisition of word semantics by a humanoid robot via interaction with a human tutor. In *New Frontiers in Human-Robot Interaction*, K. Dautenhahn and J. Saunders (Eds.). John Benjamins Publishing Company, Amsterdam, 211–234.
[41] John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge.
[42] Jeffrey M. Siskind. 2001. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *J. Artific. Intell. Res.* 15 (2001), 31–90.
[43] Catherine E. Snow. 1977. Mothers' speech research: From input to interaction. In *Talking to Children: Language Input and Acquisition*, Catherine E. Snow and Charles A. Ferguson (Eds.). Cambridge University Press, Cambridge, England, 31–49.
[44] René A. Spitz. 1957. *No and Yes: On the Genesis of Human Communication*. International Universities Press, New York.
[45] Paul Spivey. 2007. Getting in Trouble. Retrieved from https://youtu.be/QEGZ4RR6mZ8.
[46] Luc Steels and Jean-Christophe Baillie. 2003. Shared grounding of event descriptions by autonomous robots. *Robot. Auton. Syst.* 43 (2003), 163–173.
[47] Karla Štěpánová, Frederico B. Klein, Angelo Cangelosi, and Michal Vavrečka. 2018. Mapping language to vision in a real-world robotic scenario. *IEEE Trans. Cogn. Dev. Syst.* 10, 3 (2018), 784–794.
[48] Francesca Stramandinoli, Davide Marocco, and Angelo Cangelosi. 2017. Making sense of words: A robotic model for language abstraction. *Auton. Robots* 41, 2 (2017), 367–383.
[49] Yuuya Sugita and Jun Tani. 2005. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adapt. Behav.* 13, 1 (2005), 33–52.
[50] TiMBL. 2012. Retrieved from http://ilk.uvt.nl/mblp/.
[51] Michael Tomasello. 2003. *Constructing a Language*. Harvard University Press, Cambridge, MA.
[52] Francisco J. Varela, Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA.
[53] Virginia Volterra and Francesco Antinucci. 1979. Negation in child language: A pragmatic study. In *Developmental Pragmatics*, Elinor Ochs (Ed.). Academic Press, New York.
[54] Chen Yu, Linda B. Smith, Krystal A. Klein, and Richard M. Shiffrin. 2007. Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same? In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, New York, 737–742.

Received August 2018; revised June 2019; accepted August 2019