Abschlussarbeit zur Erlangung des akademischen Grades eines Bachelor der Naturwissenschaften (BSc.)

Eine experimentelle Studie zur Vorstellung eines neuen standardisierten Stimulus- Sets zur Erfassung des Erkennens von Hilfsbedürftigkeit bei Kindern (NeoHelp)

An experimental study introducing a new standardized stimulus set for studying need-of-help recognition in children (NeoHelp)

vorgelegt von

Aenne A. Brielmann

Matrikelnummer: 01/755188

an der

Universität Konstanz

Fachbereich Psychologie

Betreuer: Prof. Dr. Thomas Götz
Table of contents

Danksagungen ............................................................................................................................... 3
Zusammenfassung ......................................................................................................................... 4
Abstract ........................................................................................................................................ 6
Introduction .................................................................................................................................. 7
Methods ....................................................................................................................................... 12
  Ethics statement ....................................................................................................................... 12
  Stimuli ...................................................................................................................................... 12
Technical apparatus and stimulus presentation ........................................................................... 15
Experimental design ................................................................................................................ 17
Experimental procedure ......................................................................................................... 19
Data collection ......................................................................................................................... 21
Setting ...................................................................................................................................... 23
Participants and exclusion criteria ............................................................................................ 24
Analysis strategy ..................................................................................................................... 26
Results ......................................................................................................................................... 29
  Picture similarity ..................................................................................................................... 29
  Description of the participant population ............................................................................... 29
  Comparison of picture variations within situations ............................................................... 30
Differences in response characteristics between the two picture-size subgroups .................... 33
Changes in hit rates and RTs according to task demands ........................................................... 34
Ambiguous situations ............................................................................................................... 38
  P3-help-side: open-ended direct no-NoH / NoH comparison ............................................. 39
    Data from large -picture subgroup ....................................................................................... 39
    Data from small-picture subgroup ...................................................................................... 41
  P2-help: NoH-recognition after short stimulus presentation .............................................. 42
    Data from large -picture subgroup ....................................................................................... 42
    Data from small-picture subgroup ...................................................................................... 42
  P1-bird: human-bird-distinction after short stimulus presentation .................................... 43
    Data from large -picture subgroup ....................................................................................... 43
Data from small-picture subgroup ......................................................................................44
Summary: Suitability of the stimulus set for investigating responses on different tasks ......44
Changes in hit rates and RTs according to stimulus content ..............................................45
   P1-bird: human-bird-distinction after short stimulus presentation .................................46
   P2-help: NoH-recognition after short stimulus presentation ..........................................51
   P3-help-side: direct no- NoH / NoH comparison without time restriction ....................58
Changes in hit rates and RTs according to scenario sequence .......................................65
   Data from large-picture subgroup ..............................................................................65
   Data from small-picture subgroup ..............................................................................68
   Summary......................................................................................................................68
Discussion ......................................................................................................................70
References ......................................................................................................................77
Appendix .........................................................................................................................81
Danksagungen

Ich möchte als Erstes Prof. Thomas Götz und Dr. Margarita Stolarova für die Betreuung dieser Abschlussarbeit danken. Mein besonderer Dank gilt Dr. Margarita Stolarova, ohne deren Idee, ein neues Stimulus Set wie das hier vorgestellte zu schaffen, diese Arbeit nie zustande gekommen wäre und deren Engagement, Geduld und Begeisterungsfähigkeit die Arbeit an diesem Projekt zu einer prägenden positiven Erfahrung für mich gemacht haben. Mein Dank gebührt auch allen Kindern, die an dieser Studie teilgenommen haben, sowie allen an der Datenerhebung beteiligten Wissenschaftlern und Helfern: Tobias Heikamp, Irina Iljuschin, Muna Pohl, Monique Orzechowski, Corinna Wolf und Stefan Wolf.
Zusammenfassung

werden Standardwerte für Genauigkeit und Antwortzeiten als Vergleichswerte und zur Auswahl von Bildern für Folgestudien bereitgestellt.
Abstract

It is a general concern in experimental research to ensure that results do not relate to systematic perceptual differences when comparing responses to stimulus categories of different content. Confounding effects can be prevented by using standardized stimuli and carefully controlling their perceptual properties. The study of socioemotional development so far lacks such stimulus sets. Need-of-help recognition is one important aspect of socioemotional development and is a necessary precondition for active helping. In this thesis, I present the NeoHelp stimulus set that has been created to allow standardized testing of need-of-help recognition with clinical and normative populations of different ages. At this point in time the NeoHelp stimulus set consists of 82 black-and-white comic-like drawings of 15 different everyday situations. These pictures can be arranged into 41 pairs depicting the same agent as either in need of help or not. At the same time control picture pairs can be arranged showing a human or a bird in the same situation. Thus experimental paradigms assessing need-of-help recognition can be implemented using the identical stimulus material as used for implementing control paradigms assessing a basic-level human-animal distinction. In this thesis, I provide evidence that the picture pairs of the NeoHelp are very similar regarding low-level picture properties. In addition I present empirical data obtained from a sample of 80 children with a broad age range (3 to 13 years) for three different mutually controlling and complementing paradigms. The results demonstrate that children correctly categorize the pictures’ content regarding both tasks, need-of-help-, as well as human-bird-distinction. I also show that task requirements coherently determine which aspects of the pictures’ content influence response characteristics. Moreover, standard response characteristics (hit rates and response times) for the selection of stimuli and comparison in future studies are provided.
Introduction

It is a challenge for researchers to construct their experiments in a way that allows them to draw valid conclusions about what manipulations lead to distinct outcomes. One of the crucial steps in planning such experiments is choosing appropriate stimulus material. Scientific research therefore benefits from access to pretested standardized stimulus sets. They provide information about the stimuli’s characteristics as perceived by an average population and either specify to what extent or rule out the possibility that experimental manipulations are confounded with changes in non-considered stimulus dimensions, such as for example the perceptual properties of the stimuli. Moreover, such sets, if used repeatedly, provide directly comparable data, thus increasing knowledge over and above the insights of single experiments, e.g. [1,2].

Concerning visual stimulus material at least two important dimensions of stimulus quality need to be discussed: stimulus content (e.g. which object is depicted, to which category it belongs and how well the intended content is recognized by different participant populations), and low-level perceptual stimulus properties (such as e.g. luminance [3], color [4] and spatial frequency [5]). Usually, researchers are interested in attributing differences in outcomes to either stimulus content or low-level perceptual stimulus properties. Yet, both dimensions are often confounded and changes in one dimension can lead to unintended changes in the other. Detailed information about the properties of each stimulus is necessary for drawing well grounded conclusions about the relationship between certain aspects of the stimulus material and the effects observed. It is therefore advisable to carefully consider all possible confounds between low-level stimulus properties and stimulus content when creating new (visual) stimuli.

The need for stimuli comparable on non-manipulated dimensions applies to a variety of research methods, both behavioral and psychophysiological. When comparing experimental
outcomes with regard to the content of stimuli, it is important to ensure that results do not relate to systematic perceptual differences between the stimulus categories, e.g. [6,7]. It has been noted specifically for ERP studies that one should “never assume that a small physical stimulus difference cannot explain an ERP effect” ([6], p. 74), and this notion can be extended to other research domains. In psychophysiological experiments, e.g. in EEG and fMRI but also eye movement recordings, picture properties, such as complexity, e.g. [8], contrast, e.g. [9,10], intensity, e.g. [11], and color (e.g. [4]; see also [12] for a review) have been shown to profoundly influence participants’ responses.

Many standardized stimulus sets exist for objects, e.g. [13–15], as well as for scenes [16,17], faces, e.g. [18], and other categories of visual stimulus material, such as mixed object/scene stimulus sets, emotion-evoking pictures and movies of human actions [19–21]. Some of them have been used for many years and have helped advance research theories, as for example on emotional perception and processing using the IAPS stimuli sets, e.g. [2,22], and methods (e.g. morphing algorithms for face pictures [23]). Standardized stimulus sets provide a valuable control over experimental manipulations. They facilitate operationalization and enhance comparability between studies.

So far, however, no standardized set of material that encompasses stimuli relevant for social behavior exists. Some recent studies investigating the development of theory of mind (ToM) have used comic stories as stimulus material in brain imaging studies [24–26]. Other picture sets have been employed for evaluating deficits in social emotional behavior, self-concepts [27,28], expression of feelings [29], autism [30], and schizophrenia [31]. Unfortunately, only limited information on perceptual properties and no normative base line data are available for these stimuli, even though at least one of them has been used repeatedly [24, 25]. Another
major limitation of existing material concerns construct validity: visual stimuli depicting social situations have scarcely been pretested with regard to clarity of illustration, easiness of categorization or general understandability.

This might be one reason why many studies of complex social behavior and their development rely on real social interactions (e.g. [32,33], see also [34] for a book chapter summarizing relevant studies). At the same time social behavior is one field of psychological research that constantly attracts attention, especially in developmental psychology. Socioemotional abilities have been shown to undergo constant development throughout childhood, well into adolescence (e.g. involving 5 to 14 year-olds: [35]; for a review considering ages 3-18: [36,37]; for a longitudinal study see [38]). Active helping, one kind of prosocial action and one possible and desirable consequence of need-of-help recognition, has been assessed extensively in studies relying on real-life social interactions (e.g. [32,33], see also [39] for a book chapter summarizing relevant studies). In recent years, developmental psychology has seen a rise in the assessment of active helping in infants and toddlers documenting children’s early willingness and ability to help [40–45]. While it can be assumed that when helping occurs, the need-of-help has been recognized, the reverse inference cannot be made: There are many reasons not to help and one of them is not to have realized that someone needed help. Assessing when and under what circumstances humans are able to identify need-of-help will increase our understanding of social behaviors, the ontogeny of helping and potentially of factors endangering prosocial development. Therefore it was the aim of my thesis to develop a standardized set of visual stimuli, overcoming the methodological limitations mentioned above.

Investigating need-of-help recognition is thus a worthy research path, that has so far received little attention, possibly due in part to lack of successful experimental
operationalizations and accessible stimulus material. Therefore my thesis work was to create pictures of situations understandable for children of different ages as well as for adults and clinical populations, i.e. simple comic-like black and white line drawings. These drawings depict 15 harmless and well known everyday situations. Variations of these situations, either depicting somebody in need-of-help or not, were created in a way ensuring maximum low-level perceptual similarity. In order to enable application of a control task that requires only basic human-animal distinction, picture variations showing a bird in the same situation as the human were created again maintaining maximum perceptual similarity. This control task was employed in order to enable unraveling general categorization abilities from specific need-of-help recognition abilities. This is especially important, when e.g. for economic reasons perceptual and differentiation abilities cannot be assessed in depth, but it is still highly relevant, whether general developmental differences or specific differences regarding need-of-help recognition are observed.

It is the main purpose of this thesis to determine whether the stimuli employed fulfill perceptual and content requirements necessary, i.e. perceptual similarity of the stimuli as well as general clarity of content according to both, species (here: human vs. bird) and need-of-help-related categories. Thus, it concerns the NeoHelp-stimuli, their perceptual properties and their suitability for use in empirical studies with children. This thesis provides a basis for future research investigating the influences of child-related factors, such as age and gender, but therefore cannot take these factors into consideration itself. Evidence is provided that the NeoHelp stimuli are highly similar regarding low-level picture properties and thus are unlikely to cause confounding effects of content and low-level perceptual stimulus properties. In order to test the recognizability of the newly created NeoHelp, response characteristics from 80 children
(3 to 13 years) in three different mutually controlling experimental paradigms were obtained. This data provides baseline values for response accuracy (hit rates) and response times (RTs) for the NeoHelp stimulus set (see appendix). The results show that the NeoHelp stimuli are reliably categorized both according to help-content as well as to species categories. They demonstrate that the need-of-help recognition and control human–bird differentiation tasks yield distinct response and effect patterns.

In conclusion, the NeoHelp stimulus set fulfills the requirement of standardized visual stimuli and will aid research on need-of-help recognition in the broader domain of socioemotional development.
Methods

Ethics statement

Parents gave written informed consent according to the principles of the Declaration of Helsinki before their children participated in this study. Special care was taken to ensure that parents and children understood that their participation is voluntary and can be ended at any time without causing disadvantages to the participants.

Stimuli

The stimulus set presented here was created in order to allow standardized behavioral and psychophysiological testing of need-of-help (NoH) recognition in children. The NeoHelp stimulus set currently consists of 82 black-and-white drawings (600px x 800px) that show children of different ages and sexes in 15 individual everyday situations. A reference picture for each situation was hand-drawn and converted to a black and white vector graphic using Adobe Illustrator CS 4. This picture format allows altering only distinct lines of a picture in a controlled way. Applying this technique enabled the creation of perceptually highly similar picture variations, most importantly picture pairs depicting need-of-help (NoH) and no-need-of-help (no-NoH). Within each NoH-picture pair differences only concerned those aspects of the picture that indicated NoH (see Fig.1 for an example situation; see appendix for details of all situations).
Figure 1. Example situation and picture generation process. The situation “blocks” with all its variations is shown here as an exemplarily. Pictures framed in purple illustrate the need-of-help variations; pictures framed in green the no-need-of-help variations. Solid arrows indicate the picture from which variations were derived by changing as little lines as possible. Dashed arrows indicate which pictures were combined to generate variations in no-need-of-help pictures (here a bird variation and a variation of gender): Single features were transferred from the picture variation to the no-need-of-help reference picture.
Control pictures of birds were created in the same way. The bird control stimuli were intended to enable implementation of a basic human-bird distinction, which allows disentangling general effects of categorization abilities from specific help-content related effects. Birds were shown in the same situations as humans to ensure maximum low-level picture similarity between the pictures of any human-bird pair. This similarity was achieved by altering only those lines of the reference picture necessary to generate the corresponding picture variation. In order to increase the number of trials per situation for the human pictures, additional variations for most situations (10 out of 15) with humans, i.e. nine times of gender (showing a boy and a girl), once of age (toddler and older child) and once of ethnicity (African and Western European decent) were derived. Perceptual similarity was ensured by altering only distinct lines of the reference picture to generate the corresponding picture variation. Resulting changes in NoH-picture features were then transferred to the original no-NoH picture. Figure 1 shows all pictures for one example situation and how variations were successively derived from the reference picture.

Table 1 lists all 15 situations with their variations. A complete listing of all pictures, including the picture itself and its properties, is provided in the appendix. In sum there were 41 different pictures pairs (NoH – no-NoH) for 15 situations, 26 depicting children and 15 control picture pairs depicting birds in the same 15 situations. However, the NeoHelp stimulus set can easily be expanded by additional situations or picture variations.
Table 1. List of situations in the stimulus set, their variations and number of pictures

<table>
<thead>
<tr>
<th>Situation</th>
<th>NoH depiction</th>
<th>No-NoH depiction</th>
<th>Variations derived</th>
<th>Number of Pictures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>Try to put the last block on top of a tower</td>
<td>Build a tower with blocks</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Boat</td>
<td>Try to reach sth. in the water</td>
<td>Catch sth. out of the water</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Branch</td>
<td>Be stuck on a twig</td>
<td>Walk past a branch</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Climb</td>
<td>Try to climb a table</td>
<td>Stand in front of a table</td>
<td>Bird</td>
<td>4</td>
</tr>
<tr>
<td>Door</td>
<td>Try to reach door knob</td>
<td>Open a door</td>
<td>Bird</td>
<td>4</td>
</tr>
<tr>
<td>Gap</td>
<td>Try to cross a large gap</td>
<td>Step over a small gap</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Shelf</td>
<td>Try to reach a ball inside a shelf</td>
<td>Grasp a ball inside a shelf</td>
<td>Bird, ethnicity</td>
<td>6</td>
</tr>
<tr>
<td>Shirt</td>
<td>Try to take off shirt</td>
<td>Stand</td>
<td>Bird</td>
<td>4</td>
</tr>
<tr>
<td>Sit</td>
<td>Reach out to be lifted up</td>
<td>Sit on the floor</td>
<td>Bird</td>
<td>4</td>
</tr>
<tr>
<td>Stair</td>
<td>Try to reach top of stair step</td>
<td>Climb up a step of a stair</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Table</td>
<td>Try to reach ball on table</td>
<td>Grab ball from table</td>
<td>Bird</td>
<td>4</td>
</tr>
<tr>
<td>Table_chair</td>
<td>Try to reach a ball on a table</td>
<td>Grasp a ball on a table</td>
<td>Bird, age</td>
<td>8</td>
</tr>
<tr>
<td>Apple*</td>
<td>Try to reach an apple on a tree</td>
<td>Pick an apple</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Bucket*</td>
<td>Try to lift a heavy bucket</td>
<td>Put sth. into a bucket</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Drawer*</td>
<td>Try to open a drawer</td>
<td>Open a drawer</td>
<td>Bird, gender</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>82</td>
</tr>
</tbody>
</table>

Note. Variations refer to: gender = boy and girl; ethnicity = African and Western European decent; age = kinder garden age and toddler. Note that the age variation of “Table_chair” also had a separate bird control picture pair. *situations that were only correctly displayed in smaller size

Technical apparatus and stimulus presentation

Stimuli were presented on the screens of five different laptops using Presentation (version 16.0) software. In order to make it easier for children of different ages to handle the tasks and to standardize the response collection, five identical regular keyboards were adapted for the use as response devices. All keys were covered with a cupboard contraption, except for the two response keys, which were laminated and color coded according to the experimental presentation (see Fig. 2). Color codes were counter balanced for left and right responses across different PCs.
Figure 2. Experimental setting. The stand was located outside and was part of a child-directed family fair. A maximum of five children could participate at a given time. One laptop collecting responses and showing stimuli on its screen, one keyboard (with color-coded response keys and a cupboard covering task-irrelevant keys) and a set of headphones were assigned to each child.

On four out of the five screens pictures were presented in larger size (max. visual angle = 15.62° picture height), on one laptop pictures were significantly smaller in size (max. visual angle = 11.75° picture height). Children were assigned to computers randomly. Due to restrictions of screen resolution and size, pictures of three situations (“apple”, “bucket” and “drawer”) were only analyzed if acquired on the laptop with smaller picture presentation. Thus, data from 12 different situations was analyzed if obtained from the four laptops showing larger pictures and data from all 15 situations was analyzed if obtained from the laptop showing small pictures.
Experimental design

The experimental design consisted of three separate paradigms united through a child directed cover story: P1-bird, P2-help and P3-help-side (see Fig. 3 for a detailed summary). Responses were always made by pressing one of two color coded keys on the response keyboard. For all paradigms, the order of picture presentation was random, as was the selection of the pictures for the training phase. Children were always instructed to respond as fast and as accurately as possible. A child directed cover story motivated children to do so (see “Experimental procedure” for a detailed description).

P1-bird- and P2-help had identical two-alternatives-forced-choice (2AFC) task setups and differed only in terms of instructions. In the control task P1-bird children were asked after stimulus off-set to indicate whether they had just seen a bird or a human. In P2-help the identical stimulus presentation was followed by the question of whether children saw someone (bird or human) in need-of-help (NoH) or not. Each trial of the two paradigms P1-bird- and P2-help consisted of a 100 ms inter stimulus interval (ITI), 500 ms stimulus presentation and a decision screen, where participants were asked to make a 2AFC-response (see Fig. 3). Paradigm P3-help-side was a pair-wise picture-selection-task without time restriction. Trials were preceded again by an ITI of 100 ms. Afterwards a no-NoH/NoH pair (humans or birds) was presented on the screen until the child made a response. Children were asked to indicate which one of the two pictures showed someone in need for help.
Figure 3. Experimental design-variations. The overall paradigm sequence shown on the left side was the same for all three paradigms. The right side of the figure illustrates trial sequences, which were always the same for training and experimental trials. Note that P1-bird and P2-help had identical time restricted stimulus presentations (500 ms), trial sequence, randomization routines and experimental designs (2AFC); decision and response took place after stimulus offset. Only introduction and instruction differed between P1-bird and P2-help: in P1-bird children were asked to indicate whether the picture showed a human or a bird, while in P2-help they were asked whether the picture showed someone in need-of-help (NoH) or no-NoH. In P3-help-side on the other hand, stimuli were presented pair-wise, presentation was continued as long as needed for the child to respond, and thus decision was taken during the stimulus presentation.
Paradigms were designed to allow differential assessment of which task demands caused distinct behavioral effects: Comparing results from P2-help and P1-bird disentangled influences of basic categorization abilities (and thus of general developmental characteristics such as speed of processing) from those of stimulus presentation and response mode. Comparing P2-help and P3-help-side disentangled influences of task load through time restriction from those of the NoH-recognition. Thus, any effect that emerges in both help-related paradigms (P2-help and P3-help-side) and not in P1-bird can be attributable to the need-of-help recognition task demand.

The order of paradigm P1-bird and P2-help presentation differed depending on the different PCs. The assignment of children to PCs insured random participation. P3-help-side was always conducted last, since the kind and duration of presentation made familiarization with the stimuli likely. This procedure resulted in two randomly assigned order variations: 1) P1-bird, P2-help and P3-help-side (N = 39) and 2) P2-help, P1-bird and P3-help-side (N = 31).

**Experimental procedure**

All three paradigms started with introduction screens explaining the cover story, showing the main figure “Blobs” and his alien family. Further screens prompted responses regarding children’s participant number, their age and whether or not they had any pets (see Fig. 3). These introductions were always directed by an experimenter, who typed the responses and administered the training trials before starting the actual experiment. All three paradigms were preceded by a minimum of three training trials, in order to ensure that the participating child had understood the instruction and was capable of operating the response buttons. If needed, the experimenter could continue with training trials, until the participants mastered the response requirements. Children were instructed to respond as fast and accurately as possible in each paradigm. They saw a motivating picture showing one of Blobs’ family members accompanied
by applause sounds after completing half of the trials. Children wore headphones during the experiment to reduce distracting noise and to be able to hear applause sounds for encouragement.

Before starting P1-bird, the experimenters explained that Blobs had just lost his pet-bird and needed help finding it. Therefore, the children needed to watch the flashing pictures carefully and indicate by pressing a button whether they saw a picture of a human, or of a bird (Blobs’ lost pet). At the end of this paradigm, children saw a happy Blobs reunited with his bird and were praised for helping him to find his pet.

Before starting P2-help, the experimenters explained that Blobs, while flying around Earth fast and catching only a glimpse of different situations, had difficulties understanding when someone on Earth needed help. Special care was taken to explain that it did not matter, whether humans or birds needed help, but that the question in general was whether anyone needed help or not. So again, in order to help Blobs, children needed to watch the fast flashing pictures very carefully and indicate after each one by pressing a designated button whether it depicted someone in need of help or not. At the end of this paradigm children were praised for doing a good job. Again, a picture of a happy and grateful Blobs was shown.

In P3-help-side, the experimenter explained to the children that Blobs was still confused about behavior on Earth and continued to have difficulties understanding if and when somebody needed help. Children had to indicate with a button press, whether the help-variation was shown on the left or the right side.

Depending mainly on the age of the children, total testing time per paradigm including instruction and training varied between 4 and 12 min. Children were free to end the session at any time. If the experimenter came to the conclusion that requirements of the task by far exceeded the children’s ability (mostly because of young age, e.g. below 4 years), they suggested
completing only P3-help-side, where pictures were presented as long as needed for the child to react.

**Data collection**

For every participant the following demographic information was obtained: birthday, age in years, gender, as well as whether or not the child owned a pet and if yes, what kind. The experimenters recorded parental interference, as well as specific difficulties resulting from the participants’ age, motivation or developmental level, if necessary. In addition, response times and key press responses were collected via the presentation software (Presentation version 16.0). Response times (RTs) for P1-bird and P2-help depictions were recorded as the time elapsed from the onset of the decision screen to the button press initiated by the child. In P3 RTs were recorded from the onset of the picture pair presentation to the button press initiated by the child.

All further data processing was done using the free GNU-software R (version 3.0.0). Trials in which stimulus presentation was followed by more than one key press were classified as “multi-responses”, excluded from analyses and replaced by missing values. Figure 4 shows detailed quantities of data loss due to data processing procedures and exclusion of participants. Note that no multi-responses were detected for P3-help-side (no time restriction of stimulus presentation). Percentages of correct responses were calculated relative to the total number of unambiguous responses (total – multi-responses). Subsequently the term hit rates will be used to refer to the percentage of correct responses.
Figure 4. Exclusion procedure and data cleansing. (N = number of different children from whom data is considered; n = number of trials). “Skipped trials” were the result of children beginning but not finishing a paradigm. “Multi-responses” refers to exclusion of all trials for which more than one key press was recorded. The “cropped situations” were “apple”, “bucket” and “drawer”; due to technical difficulties their presentation on 4 out of 5 laptops was impaired. “Participant-wise exclusion” occurred because of technical difficulties with the equipment (2), parental interference (1), because children had already done the experiment in a pretest-setting (2) or because of significant cognitive delays (1). “Wrong paradigm order” occurred if (due to experimenter’s error) children completed the P3-help-side paradigm before any of the other two paradigms. “Non-compliance” was defined as being the case when a paradigm’s mean RT was above 4 s or below 300 ms or when mean hit rate was below chance (50%). RTs were considered “unrealistic” if they were above 4 s, below 100 ms or 4 SDs apart from a child’s mean RT. The “% of valid trials” refers to the proportion of attained trials that has been used for analyses. Note that two children changed PCs, contributing data to both picture-size subgroups depending on paradigm.

Setting

Data collection for this study occurred during an open-air child directed festival on September 8th, 2012 in the German city of Konstanz. The experimental setting is illustrated in Figure 2. Four trained experimenters were present at the stand at all times. One of them talked to the parents of interested children, obtained informed written consent and recorded the children’s gender and data of birth. The other three experimenters took care of the children at the computers, explained the procedure, conducted the training trials, started the individual
paradigms in a predefined randomized order and recorded the participant numbers, as well as any peculiarities occurring during testing.

Every child who expressed interest to participate and whose parents agreed to this in writing received a “research passport”. This passport was a card with the child’s given name, participant number and pictures symbolizing the three experimental tasks. When children completed one of the tasks, they received a stamp. After completing all three tasks, or when they lost interest in continuing, children received a small present and kept their research passports.

**Participants and exclusion criteria**

No child was refused participation; intentional pre-selection of any kind did not take place. At least one paradigm was begun by 89 German speaking participants (mean age = 7.91 yrs., SD = 2.52, 54 boys), during the course of the whole day starting at 11 a.m. and finishing at 5 p.m. It can be assumed that the participating children either lived in one of the two bordering towns Konstanz (in Germany) and Kreuzlingen (in Switzerland) or their surroundings.

A summary of the exclusion criteria and the complete data cleansing procedure, including the number of trials for each of the three paradigms at any given step, is provided in Figure 4. Data of 80 different children (50 boys) between the ages three and 13 (mean age = 8.19 yrs, SD = 2.26) was included in further analyses for at least one paradigm. The majority of them (60 children, 37 boys) completed a variation of the experiments with large picture presentation. I will subsequently refer to those children as “large-picture subgroup”. The 22 children (14 boys) who used the fifth computer (through random assignment) saw a variation with smaller pictures and will be referred to as the “small-picture subgroup”. Note that one girl and one boy changed PCs between paradigms because of technical difficulties, thus contributing data to both picture-size subgroups for different paradigms.
The total number of children included for analysis, distribution of gender and pet-ownership for the total sample, as well as the picture-size-subgroups are presented in Table 2. The age distribution for each gender is illustrated in Figure 5. Because of the paradigm-wise exclusion criteria and because not all children completed all three paradigms, data from slightly different subsamples was available for analysis of each paradigm. Details regarding concrete trial numbers are provided at the beginning of each analysis.

Table 2. Distribution of gender, age and pet-ownership for the sample and picture-size subgroups

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of Children</th>
<th>Gender</th>
<th>Pet</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Yes</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>62.50 %</td>
<td>37.50 %</td>
<td>33.75 %</td>
</tr>
<tr>
<td>Large Pictures</td>
<td>60</td>
<td>59.00 %</td>
<td>41.00 %</td>
<td>34.00 %</td>
</tr>
<tr>
<td>Small Pictures</td>
<td>22</td>
<td>61.67 %</td>
<td>38.33 %</td>
<td>35.00 %</td>
</tr>
</tbody>
</table>

*Note.* Two children changed PCs, resulting in descriptive data for both picture-size subgroups

*Figure 5.* Number of children whose data was used for analyses. Red proportions of bars represent the number of girls, blue proportions the number of boys.
**Analysis strategy**

All statistical analyses were conducted using the GNU-software R (version 3.0.0). The analyses conducted here focused on assessing whether the newly created stimulus set is suitable for assessing NoH-recognition in children. Special emphasis was put on detection of unsuitable stimulus material and on exploring which picture properties influence response characteristics in NoH-recognition tasks.

To evaluate the similarity between the pictures comprising a NoH-/no-NoH pair as well as between the variations of NoH-depictions, the structural similarity index (SSIM) between those pictures was calculated. SSIM was proposed as a measure of picture quality by quantifying the visibility of errors in a distorted picture. It is employed as a measure of picture similarity here, because the comparisons made refer to differences between pictures in general and, as the authors put it, “can be thought of as a similarity measure for comparing any two signals” [46].

In order to provide information about the standardization population, descriptive statistics were obtained. Two sample subgroups were created on the basis of whether the children saw the larger or smaller pictures. For the whole sample (n = 80) as well as both subgroups, Shapiro-Wilk-Tests for normal distribution of age, as well as exact binomial tests (π = .50) on distribution of gender and pet-ownership were conducted. Systematic differences in age distribution between picture-size subgroups were explored using a Wilcoxon rank sum test. Systematic differences in the dichotomous gender and pet-ownership variables were explored using exact binomial tests (π = estimated proportion of hits in the other subgroup).

It was first analyzed whether it is acceptable to cluster picture variations according to the situation they depicted rather than to analyze each picture pair independently. Therefore, mean hit rates as well as RTs were compared between human picture variations of those 10 situations
for which variations (gender, age or race) were available. Data from bird depictions was excluded from this analysis, as there were no equivalent variations for bird pictures. Thus, 12 separate 10 x 2 (situation x picture-variation) ANOVAs were conducted for each paradigm (3) and each picture-size subgroup (2) for hit rates (6) as well as RTs (6).

Second, influences of picture-size on response characteristics were investigated in order to determine whether this factor has to be considered when using the stimulus set in different tasks. Differences in overall hit rates and RTs between picture-size subgroups were assessed on the basis of data from all 12 situations that were properly shown in both sizes, using 12 x 2 (situation x picture-size) ANOVAs for each paradigm.

Third, differences between paradigms regarding response characteristics (RTs and hit rates) were assessed using N x 3 (situation x paradigm) ANOVAs for each picture-size subgroup, with N being the number of situations analyzed in each subgroup.

Fourth, in order to determine whether the situations included in the stimulus set enabled children to distinguish between need-for-help (NoH) and no-NoH-pictures in P2- and P3-paradigm, exact binominal tests (π = .50) were conducted on hit rates per situation and across all situations. Situations for which the estimated hit rate did not exceed chance were classified as “ambiguous”. Analogue exact binominal tests were conducted for P1 in order to determine whether pictures of birds and humans were clearly discernible by children. Only unambiguously classifiable situations were included in further analyses of mean hit rates and RTs.

Fifth, the influence of specific picture properties on children’s responses was assessed by means of four different factors: situation, need-of-help-depiction (NoH-depiction), need-of-help-side (NoH-side) and bird-depiction. Note that only one of the two factors NoH-depiction and NoH-side can be considered in each analysis, as NoH-depiction only applies to the paradigms
P1-bird- and P2-help, while NoH-side only applies to P3-help-side. Three-way N x 2 x 2 ANOVAs, with N being the number of unambiguous situations included, were conducted for each paradigm and picture-size subgroup. The first factor was always situation. The second factor was always bird-depiction. The third factor was either NoH-depiction (for P1 and P2) or NoH-side (for P3). A total of 12 N x 2 x 2 ANOVAs were calculated for mean hit rates (6) and mean RTs (6) for each picture-size subgroup (2) and each paradigm (3) separately.

Regarding mean RTs, additional analysis were conducted to investigate whether RTs were different for correct vs. incorrect responses. Therefore, N x 2 (situation x correctness) ANOVAs were computed for each paradigm and picture-size subgroup separately, with N being the number of situations available for analysis. The relationship between hit rates and RTs of responses to different situations was clarified by testing significance of Pearson’s correlation coefficients.

The influence of the order in which paradigms were absolved was explored using N x 2 (situation x paradigm sequence) ANOVAs, with N being the number of situations analyzed in the given picture-size subgroup. Only data for unambiguous situations was considered. Also, only data of children who absolved all three paradigms in one of the two planned sequences is included here (N = 66), as the number of trials for the unplanned paradigm sequences was too low. Separate ANOVAs were calculated for each paradigm and each picture-size subgroup.

Further exploration of significant interactions revealed by ANOVAs was done computing Tukey honest significant difference (HSD) tests.
Results

Picture similarity

The structural similarity index (SSIM, originally described in [46]) for NoH and no-NoH picture pairs ranged from 73% to 91% for human depictions, showing that objective picture properties were similar within these picture pairs. For bird depictions SSIM comparing NoH- and no-NoH variation of bird pictures ranged from 68% to 95%, but was below 83% only one time for the situation “shelf”. Also, bird and human depictions were perceptually highly similar, as indicated by a SSIM ranging from 81% to 94%. What is more, the low SSIM for comparison between NoH-pictures and their variations demonstrates that objective similarity is also high within each situation: With one exception (the variation of ethnicity, SSIM = 65%), SSIM for human variations ranged from 78% to 99%. Thus, low-level stimulus properties are unlikely to substantially influence responses to corresponding pictures in the NeoHelp stimulus set. Detailed SSIMs for each situation are listed in the appendix.

Description of the participant population

This first standardization population for the NeoHelp stimulus set included 80 children aged from three to 13 years. Age was normally distributed across participants, more boys than girls participated and the majority of children did not own a pet. Population characteristics (age, gender and pet-ownership) were not different between picture-size subgroups. Detailed descriptive statistics and comparisons of subgroups are provided below.

First, age was normally distributed, $W = 0.86$, $p = .06$ (see Fig. 5 for details of the age distribution). This was also true for the small-picture subgroup ($n = 22$), $W = 0.88$, $p = .21$, but not for the large-picture subgroup ($n = 60$), $W = 0.72$, $p < .01$ (see Fig. 5 for percentages across picture-size-subgroups). Nonetheless, age distribution was overall not significantly different between picture-size subgroups, $W = 55$, $p = .05$. 
Second, the fact that the absolute number of boys was higher in the whole sample as well as both subgroups (see Table 2), lead to significant deviations of gender distribution from equal distribution in the complete sample, \( p = .03 \). This was not the case in the small-picture subgroup, \( p = .29 \), and only marginally evident in the large-picture subgroup, \( p = .09 \). Moreover, additional testing of whether gender distributions of either subgroup differed from the estimated distribution in the other proved that gender distribution was the same in both picture-size subgroups, both \( p \geq .38 \).

Third, in the complete sample as well as in the large-picture subgroup, there were significantly more children that did not own a pet than such who did, \( p < .01 \) and \( p = .03 \), respectively. Only in the small-picture subgroup pet-ownership was equally distributed, \( p = .13 \). As percentages of children owning and not owning a pet are almost identical in both subgroups (compare rows of the last columns of Table 2), it can be assumed that the diverging results for the small-picture subgroup are due to the smaller sample size (\( N = 22 \)). As for gender distribution, binomial tests on whether pet-ownership distribution in one subgroup diverged from estimated distribution in the other subgroup revealed that there were no differences between subgroups, both \( p \geq .65 \).

**Comparison of picture variations within situations**

The NeoHelp stimulus set at the moment consists of 15 different everyday situations. For each of them picture pairs, comprising of a need-of-help (NoH) and a no-need-of-help (no-NoH) depiction, were created. A picture pair either depicted a human child or a bird. For ten of these situations, additional variation pairs of the human depictions (altering gender or age or race) were derived. As these variations theoretically might contribute to differences of compared means, I tested whether responses to variations of a given situation differed significantly from
each other. Table 3, 4 and 5 list the detailed results of the ANOVAs conducted for each paradigm. Regardless of picture-size subgroup and paradigm, picture-variation within a situation never had an effect on either hit rates or RTs and also never interacted with situation. Thus, it is conceivable that picture variations do not constitute separate measures and did not distinguish between variations of human depictions in subsequent analyses. Note that therefore the results reported for each situation are derived from means across all human variations of a given situation, which increased the number of trials per situation.

Table 3. Results of ANOVAs investigating the influence of picture-variation in P1-bird

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th></th>
<th>Small-Picture Subgroup</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df Hit Rates RTs</td>
<td>df Hit Rates RTs</td>
<td>df Hit Rates RTs</td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>6 1.02 .41 1.30 .25</td>
<td>9 0.74 .68 1.29 .24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture-variation</td>
<td>1 0.04 .84 1.17 .28</td>
<td>1 0.20 .66 0.15 .70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH-depiction</td>
<td>1 4.90 .03 1.00 .32</td>
<td>1 0.27 .61 1.75 .19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x picture-variation</td>
<td>6 1.11 .35 0.19 .98</td>
<td>9 0.69 .72 0.91 .52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x NH-depiction</td>
<td>6 1.99 .06 2.48 .02</td>
<td>9 1.27 .25 1.16 .32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture-variation x NH-</td>
<td>1 0.18 .67 0.00 .96</td>
<td>1 0.32 .57 0.18 .67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>depiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x picture-variation x</td>
<td>6 0.92 .48 0.28 .95</td>
<td>9 0.99 .45 0.88 .55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH-depiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Results of ANOVAs investigating the influence of *picture-variation* in P2-help

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th>Small-Picture Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rates</td>
<td>RTs</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>F</td>
</tr>
<tr>
<td>Situation</td>
<td>6</td>
<td>6.80***</td>
</tr>
<tr>
<td>Picture-variation</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>NH-depiction</td>
<td>1</td>
<td>17.44***</td>
</tr>
<tr>
<td>Situation x picture-variation</td>
<td>6</td>
<td>0.89</td>
</tr>
<tr>
<td>Situation x NH-depiction</td>
<td>6</td>
<td>1.35</td>
</tr>
<tr>
<td>Picture-variation x NH-depiction</td>
<td>1</td>
<td>0.62</td>
</tr>
<tr>
<td>Situation x picture-variation x NH-depiction</td>
<td>6</td>
<td>0.69</td>
</tr>
</tbody>
</table>

*Note.* ***p < .001; **p < .01; *p < .05

Table 5. Results of ANOVAs investigating the influence of *picture-variation* in P3-help-side

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th>Small-Picture Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rates</td>
<td>RTs</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>F</td>
</tr>
<tr>
<td>Situation</td>
<td>6</td>
<td>6.36***</td>
</tr>
<tr>
<td>Picture-variation</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>NH-side</td>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>Situation x picture-variation</td>
<td>6</td>
<td>0.29</td>
</tr>
<tr>
<td>Situation x NH-side</td>
<td>6</td>
<td>1.12</td>
</tr>
<tr>
<td>Picture-variation x NH-side</td>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td>Situation x picture-variation x NH-side</td>
<td>6</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*Note.* ***p < .001
Differences in response characteristics between the two picture-size subgroups

In a further step of analyses, investigation of whether responses (hit rates and RTs) differed for the two picture-size subgroups took place. As task demands were expected to influence response characteristics, too, analyses were conducted for each paradigm separately. The three situations (“apple”, “bucket” and “drawer”) that were not available for analyses in both picture sizes were excluded from direct comparison across picture-size subgroups. Thus, 2279, 3947 and 4277 trials remained for analysis of data from P3-help, P2-NoH-side and P1-bird respectively.

First, differences in hit rates between picture-size subgroups were investigated: In P3-help-side, where children were allowed to compare two pictures as long as they wanted in order to decide which one shows need-of-help (NoH), hit rates were not influenced by picture-size, $F(1, 2255) = 0.02, p = .88$. The only influence on hit rates was exerted by situation $F(11, 2255) = 17.40, p < .001$. The interaction of both factors was non-significant, $F(11, 2255) = 1.22, p = .27$.

For P2-help, where pictures were presented for 500 ms only and the decision response was made after picture off-set, hit rates were significantly higher in the large-picture subgroup than in the small-picture subgroup, $F(1, 3923) = 4.40, p = .04$, MD = 3.24%, 95% CI[6.28%, 0.21%]. Situation also affected hit rates, $F(11, 3923) = 9.14, p < .001$, but again there was no interaction, $F(11, 3923) = 0.32, p = .98$.

In P1-bird, pictures were again presented for only 500 ms and the response occurred after stimulus off-set, just as in P2-help. Critically children were instructed to distinguish between pictures of birds and humans, here. Hit rates were generally higher in the large-picture subgroup, $F(1, 4253) = 51.96, p < .001$, MD = 6.87%, 95% CI[8.73%, 5.00%], as reported above for the other fast-response paradigm P2-help. In contrast to both paradigms involving NoH-recognition, hit rates did not change across situations, $F(11,4253) = 0.92$, $p = .52$, when only a distinction
between humans and birds was required. As in all other paradigms, the interaction of *situation* and *picture-size* was non-significant, $F(11, 4253) = 0.66, \ p = .77$.

Second, RTs were analyzed for each paradigm and picture size subgroup, yielding a different pattern of results compared to hit rates: In both NoH-recognition paradigms, P3-help-side and P2-help, RTs were shorter if pictures were presented in a smaller size, $F(1, 2255) = 10.63, \ p < .01, \ MD = 116 \ ms$, and $F(1, 3923) = 18.09, \ p < .001, \ MD = 124 \ ms$, respectively. What is more, RTs were also different across situations, $F(11, 2255) = 8.98, \ p < .001$ and $F(11, 3923) = 4.59, \ p < .001$ respectively. As reported for hit rates, *situation* and *picture-size* did not interact, $F(11, 255) = 1.14, \ p = .32$ and $F(11, 3923) = 0.33, \ p = .98$, respectively.

In contrast to both help-considering paradigms, RTs in P1-bird were only influenced by *situation*, $F(11, 4253) = 1.85, \ p = .04$, not by *picture-size*, $F(1, 4253) = 2.55, \ p = .11$, or the interaction, $F(11, 4253) = 1.48, \ p = .13$.

In summary, variations in picture-size altered response characteristics in a systematic way according to task demands. *Picture-size* influenced mean hit rates only if picture presentation was limited to 500 ms. In this case, hit rates were higher for larger pictures. RTs were shorter for smaller picture-sizes only if the task required children to distinguish between NoH and no-NoH depictions, not between birds and humans. Because of these systematic differential effects of *picture-size*, all following analyses will be conducted separately for small- and large picture subgroups.

**Changes in hit rates and RTs according to task demands**

To investigate whether differences in task demands result in changes in response characteristics, hit rates and mean RTs were compared across paradigms. The suitability of P3-help-side and P1-bird to serve as control variations for P2-help was thus indirectly assessed.
Table 6. Results of ANOVAs investigating the influence of \textit{paradigm} on hit rates and RTs

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th></th>
<th>Small-Picture Subgroup</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rates</td>
<td>RTs</td>
<td>Hit Rates</td>
<td>RTs</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>F</td>
<td>p</td>
<td>F</td>
</tr>
<tr>
<td>Situation</td>
<td>11</td>
<td>11.18***</td>
<td>&lt; .001</td>
<td>7.53***</td>
</tr>
<tr>
<td>Paradigm</td>
<td>2</td>
<td>176.86***</td>
<td>&lt; .001</td>
<td>1702.97***</td>
</tr>
<tr>
<td>Situation x Paradigm</td>
<td>22</td>
<td>6.29***</td>
<td>&lt; .001</td>
<td>3.18***</td>
</tr>
</tbody>
</table>

\textit{Note.} *** p < .001; **p < .01
In both picture-size subgroups, hit rates as well as RTs were considerably influenced by paradigm, situation as well as the interaction of both factors (for detailed ANOVA statistics, see Table 6). Differences of response characteristics to distinct situations are described in detail in later analyses (see section “Changes in hit rates and RTs according to stimulus content”). First, results concerning the factor paradigm will be reported.

**Figure 6.** Comparison of hit rates and response times between paradigms. Hit rates and RTs are presented across situations and for both picture-size subgroups. Light bars represent data from the small picture-size subgroup, dark bars from the large picture-size subgroup. Asterisks mark significant differences revealed by post-hoc Tukey HSD tests (***p < .001). Note that all differences between paradigms were significant but the difference in hit rates between P3-help-side and P1-bird in the small picture-size subgroup (N = 22). Error bars represent SEM.

Despite the main effects described above, situation and paradigm interacted regarding both, hit rates and RTs. Not all differences in hit rates between paradigms were significant for all situations in both picture-size subgroups. The interaction showed the same effect on RTs: Differences in RTs were non-significant for some situations. However, the same pattern revealed
by the main effect of *paradigm* (RTs for P1-bird < P2-help < P3-help-side) was present in all situations for RTs (see Figure 7). This pattern of effects emerged irrespective of picture size.

*Figure 7.* Interaction effect of *paradigm* and *situation* on response times (RTs). Mean RTs are presented for large- (circles) and small-picture subgroups (crosses) separately. Red symbols represent means for P1-bird, green ones for P2-help and blue ones for P3-help-side. Note that differences between paradigms are not significant for all situations but that post-hoc tests confirmed that RTs were significantly different between all paradigms across situations (see Figure 5 for an illustration of the main effect). Data of the situations “apple”, “bucket” and “drawer” was analyzed in the small-picture subgroup only. Error bars represent SEM.

In sum, response characteristics were profoundly and systematically influenced by task demands, as introduced by experimental designs and instructions of the three different paradigms. This was true, irrespective of the fact that responses were made to the same set of pictures and in the case of P1 and P2 in identical experimental designs except for the task
content. Lower task demands, i.e. human vs. bird distinction, lead to shorter RTs and higher hit rates. RTs were generally shorter when pictures were presented only briefly and if decisions were made after picture presentation had ended, no matter if task demands were low (human-bird-distinction) or higher (NoH-recognition). Increasing task difficulty by restricting stimulus presentation time and asking children to make a content related distinction between no-NoH and NoH-depictions decreased hit rates. When task demands were highest, i.e. picture presentation was short and the task required identifying NoH (in paradigm P2-help), hit rates were lowest and RTs were longest.

**Ambiguous situations**

After determining that task demands significantly change response patterns, individual situations were analyzed in order to investigate whether some of the situations were too ambiguous to be correctly categorized on average according to the task-related content (NoH – no-NoH in P2 and P3, human – bird in P1). The situations identified as ambiguous because of hit rates not significantly above chance were subsequently excluded from further analyses of the specific paradigm-data. Since this newly created visual stimulus set is used here for the very first time, this ambiguity-analysis is a necessary step, in order to identify situations with an unclear differentiation between NoH- and no-NoH-variations or between human and bird variations. As task demands differed between paradigms, ambiguous situations were identified for each paradigm separately. Separate analyses for each paradigm and each picture-size subgroup will be presented because of the systematic picture-size and paradigm differences reported above and summarize the results across paradigms at the end.
P3-help-side: open-ended direct no-NoH / NoH comparison

Data from large-picture subgroup

1768 trials were included for initial analysis. Dark bars of Figure 8 A show estimated probabilities of success for all 12 situations correctly shown as large picture.

Pictures of all but one situation (“climb”, estimated hit rate = 56.52%, p = .25) were identified correctly as depicting need-of-help (NoH) or no-need-of-help (no-NoH), all estimated hit rates > 78.21%, all p < .001. Responses to “climb” therefore cannot be considered to reflect a correct distinction between NoH and no-NoH depictions when decision time is not restricted and a direct comparison between the NoH and no-NoH depictions takes place. Thus “climb” was identified as ambiguous situation and excluded from further analyses of this paradigm’s data. This reduced the number of trials for analyses in the large-picture subgroup in P3-help-side from 1839 to 1676. Overall, estimated hit rate for P3-help-side across all unambiguous situations was 90.81%, p < .001, 95% CI [89.33%, 92.15%].
Figure 8. Estimated situation specific hit rates for P3-help-side (A), P2-help (B) and P1-bird (C). Estimates across variations of each situation are derived from tow-tailed exact binomial tests (π = .50). Light bars represent estimated hit rates for the small-picture subgroup, dark bars for the large-picture subgroup. Note that data from “apple”, “bucket” and “drawer” was only analyzed in the small-picture subgroup. Unambiguous situations (estimated hit rate significantly above .50) are outlined in green, ambiguous situations (estimated hit rate not significantly different from chance) are outlined in red. Error bars represent upper and lower bounds of the estimated hit rates' confidence intervals (CIs).

Overall estimated hit rates were calculated across all unambiguous situations. Lines represent overall estimated hit rates for the large- (solid) and small-picture subgroup (dashed) per paradigm. Shaded areas surrounding them indicate their CIs. Note that non-overlapping CIs in P2-help (B) and P1-bird (C) indicate significant differences in estimated hit rates between picture-size subgroups.

Data from small-picture subgroup

647 trials were included for initial analysis. Light bars of Figure 8 A show estimated probabilities of success for each situation when presented small in P3-help-side.

The pattern of effects did not differ from presentation of pictures in larger size: The only pictures not reliably categorized as NoH or no-NoH were again those of the situation “climb”, estimated hit rate = 58.62%, p = .46, all remaining estimated hit rates ≥ 77.50%, all remaining p < .001. Again, I excluded the ambiguous situation “climb” from further analyses of this paradigm’s data, as a reliable no-NoH/NoH-distinction does not seem to be possible for children
in the given task. Exclusion of “climb” reduced the number of trials for further analyses from 647 to 618. Overall, the estimated hit rate across all unambiguous situations for P3-help-side was 89.64%, p < .001, 95% CI[86.97%, 91.93%].

**P2-help: NoH- recognition after short stimulus presentation**

**Data from large-picture subgroup**

3055 trials were included for analysis. Dark bars of Figure 8 B show estimated hit rates in P2-help for each situation in large picture presentation.

Under the conditions of short picture presentation (500 ms) and responses made after stimulus off-set in P2-help, pictures of all 12 situations were on average correctly identified as showing NoH or no-NoH with a probability above chance, all estimated hit rates > 66.30%, all p < .001. Note that these 12 situations also include the situation “climb” that was ambiguous concerning NoH-depiction in P3-help-side. Overall, the estimated hit rate across all situations was 79.25%, p < .001, 95% CI [77.77%, 80.67%], which is considerably above chance but also lower compared to P3-help-side, where children directly compared pictures of no-NoH and NoH-depictions for as long as they needed.

**Data from small-picture subgroup**

1141 trials were included for initial analysis. Light bars of Figure 8 B show estimated probabilities of success for all situations if presented as small pictures.

In accordance to results obtained for the large-picture subgroup, “climb” (ambiguous in P3-help-side) was a clearly unambiguous situation for the small-picture presentation in P2-help, estimated hit rate = 75.47%, p < .001. However, the probability of pictures of the situation “gap” (not ambiguous in either P3-help-side or the large-picture sample in P2-help) to be correctly identified as NoH or no-NoH was only marginally above chance level, estimated hit
rate = 61.11%, p = .08. For all remaining situations pictures were classified as NoH or no-NoH
depiction above chance, all remaining estimated hit rates ≥ 62.50%, all remaining p ≤ .03. The
ambiguous situation “gap" was excluded from further RT and hit rate analyses of this paradigm,
reducing the number of trials from 1141 to 1069. Note however, that "gap" was only excluded
from analyses in this small-picture subgroup.

Overall estimated hit rates across all unambiguous situations was 75.49%, p < .001, 95%
CI[72.80%, 78.04%], indicating different probabilities of correctly identifying NoH-depictions
between the small- and large-picture subgroup in general, where NoH is discriminated worse if
pictures are presented in a smaller size. Moreover estimated hit rates in this paradigm, where
children had to decide from memory whether the shortly (500 ms) presented picture showed
NoH or no-NoH, are also lower compared to P3-help-side, in which children could compare
NoH and no-NoH pictures as long as they needed to make a response.

**P1-bird: human-bird-distinction after short stimulus presentation**

In P1-bird, children were asked to decide whether the picture briefly seen for 500 ms was a
human or a bird. As opposed to the paradigms considered above P1-bird did not require children
to identify need-of-help (NoH).

**Data from large-picture subgroup**

3311 trials were available for analyses. Dark bars of Figure 8 C show estimated hit rates in
P1-bird for each situation in large picture presentation. Concerning the distinction between
human and birds, there were no ambiguous situations, all estimated hit rates ≥ 90.34 %, all
p < .001. The estimated overall hit rate for P1-bird was with 94.14%, p < .001, 95% CI [93.29%,
94.92 %], the highest one of all paradigms.
**Data from small-picture subgroup**

1243 trials were available for analyses. Light bars of Figure 8 C show estimated hit rates in P1-bird for all situations in small picture presentation. Just as in the large-picture subgroup, pictures of all situations were unambiguously recognized as depicting either a bird or a human, all estimated hit rates $\geq 80.64\%$, all $p < .001$. The estimated overall hit rate for P1-bird in the small-picture subgroup was with 87.21%, $p < .001$, 95% CI [85.22%, 89.02 %], the highest one of all paradigms in this picture-size subgroup. As for P2-help the non-overlap of confidence intervals of overall estimated hit rates of small- and large-picture subgroups in P1-bird indicates that mean hit rates were generally lower in the small-picture subgroup.

**Summary: Suitability of the stimulus set for investigating responses on different tasks**

Children were capable of reliably differentiating between depictions of humans and birds for all 15 situations and regardless of picture-size subgroup (see Fig. 8 C). Pictures of almost all situations (13 out of 15, marked with green outlines in all columns of Fig. 8 A-B) of the stimulus set were unambiguously identifiable as depicting need-of-help (NoH), regardless of picture-size, presentation duration and task type (2AFC vs. picture selection). What is more, children’s responses did not differ between variations of human pictures of a given situation.

Only the situation “climb” was identified as ambiguous concerning the illustration of NoH if shown for a longer time period and if NoH- and no-NoH-depictions are compared directly (marked by red outlines in Fig. 8 A). Also, the situation “gap” has to be regarded with special care when presented for only a short time period and in large size, as response accuracy did not exceed chance level under these conditions (see red outlined bar in Fig. 8 B). However, for all other situations I reliably find above-chance NoH-recognition, no matter whether children directly compare no-NoH and NoH-pictures or whether one picture is only shortly presented.
Therefore, all situations but “climb” and “gap” are fulsome suitable for further testing which factors additionally influence NoH-recognition. Highly similar results of binomial tests of hit rates for small- and large-picture subgroups as well as for the two different paradigms involving NoH-recognition substantiate the conclusion that children perceived the difference between no-NoH- and NoH-depictions in all but two situations.

Response characteristics for the same situations were different according to the task given. Consistently, children distinguished between birds and humans more accurately and faster compared to both tasks that required a decision between NoH- and no-NoH-depiction. Hit rates were higher and RTs were longer if children were allowed to compare NoH- and no-NoH-depictions directly for as long as they needed (P3-help-side) compared to a brief (500 ms) presentation of one picture only after which children had to decide from memory (P2-help and P1-bird). The lack of overlap between estimated hit rates’ CIs of different paradigms adds more evidence to these findings (compare positions of shaded areas between Fig. 8 A-C). However it should also be noted that children did P3-help-side as the last paradigm, and consequently effects of familiarization may also have contributed to differences in response characteristics between P2-help and P3-help-side and to differences between P1-bird and P3-help-side.

**Changes in hit rates and RTs according to stimulus content**

After clarifying that picture content was perceived as intended in the vast majority of the newly created visual stimuli, I turned to investigating the influence of picture content on response characteristics. The main purpose of the following analyses is to assess the relationships between specific task requirements, stimulus content and response patterns (RTs and hit rates). Stimulus related characteristics are reflected by the factors *situation and bird-depiction*, as well as *NoH-depiction* (for P1-bird and P2-help) or *NoH-presentation side* (for P3-help-side). Additionally I explored differences in RTs for correct vs. incorrect responses (factor *correctness*).
Only data of situations that were unambiguous concerning task-related content (see section “Ambiguous situations”) was used to ensure that effect patterns were related to task fulfillment. Results of the analyses will be presented for P1-bird, P2-help and P3-help-side separately, as response characteristics have been shown to differ between paradigms (see section “Differences in response characteristics according to task demands”). For each paradigm, findings for the large-picture size subgroup will be reported first. Results obtained in the small-picture subgroup are presented in comparison to those of the large-picture subgroup.

**P1-bird: human-bird-distinction after short stimulus presentation**

*Data from large-picture subgroup*

When shown a large picture and asked to decide after off-set of a brief presentation (500 ms) whether they saw a bird or a human, children’s responses were more accurate for pictures of humans, MD = 2.35%, 95% CI[2.02%, 1.18%], p < .01. This effect was independent of whether need-of-help was depicted and of the situation presented. The detailed results of the ANOVAs for both, hit rates and RTs are given in Table 7. As response characteristics were not different across situations, or between no-NoH/NoH-depictions hit rates as well as mean RTs are shown for bird and human pictures on the left side of Figure 9. All interactions were non-significant.
Table 7. Results of ANOVAs conducted in P1-bird

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
<td>df F p</td>
</tr>
<tr>
<td>Situation</td>
<td>11 1.12 .34</td>
<td>1 1.23 .26</td>
<td>14 0.51 .93</td>
<td>14 1.43 .13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH-depiction</td>
<td>1 0.27 .60</td>
<td>1 0.11 .74</td>
<td>1 0.69 .40</td>
<td>1 0.67 .41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bird-depiction</td>
<td>1 8.29** &lt; .01</td>
<td>1 0.80 .37</td>
<td>1 5.66* .02</td>
<td>1 0.19 .66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x NH-depiction</td>
<td>11 1.14 .32</td>
<td>11 1.58 .10</td>
<td>14 0.86 .61</td>
<td>14 0.65 .82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x bird-depiction</td>
<td>1 0.73 .71</td>
<td>1 0.83 .61</td>
<td>1 0.73 .74</td>
<td>1 0.56 .90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH-side x bird-depiction</td>
<td>1 3.09 .08</td>
<td>1 1.36 .24</td>
<td>1 1.37 .24</td>
<td>1 0.02 .90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation x NH-depiction x</td>
<td>11 1.64 .08</td>
<td>11 0.95 .49</td>
<td>14 1.10 .35</td>
<td>14 0.85 .62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bird-depiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. **p < .01; *p < .05

Figure 9. Hit rates and mean RTs for human- and bird-depictions in P1-bird. Bars represent hit rates, dots mean RTs. Light bars and enclosed dots represent data from bird depictions, darker bars and enclosed dots from human depictions. Results for the large-picture subgroup are shown on the left, for the small-picture subgroup on the right. Asterisks mark significant differences in hit rates *p < .05, **p < .01, as derived from post-hoc Tukey HSD tests. Error bars represent SEM. RTs do not differ between human and bird depictions. Note that hit rates are higher for humans in the large-picture subgroup, whereas the direction of bird-depiction’s main effect was inversed in the small-picture subgroup.
The relationship between hit rates and RTs for the different situations is illustrated in Figure 10 (red circles represent data from this paradigm and picture-size subgroup). A test of Pearson's correlation coefficient showed that there was no speed-accuracy-trade-off and revealed on the contrary that in general RTs were shorter for situations that were categorized more accurately, $r = .64$, $p = .02$. RTs were not different according to any of the picture properties ($situation$, $bird-depiction$ and $NoH-depiction$) or any combination of them.

There was a main effect of correctness on RTs, $F(11, 3287) = 62.39$, $p < .001$, with correct responses being faster than incorrect ones, $\text{MD} = 252$ ms, $95\% \text{ CI}[189$ ms, $315$ ms], $p < .001$. Even though RTs were not influenced by situation directly if correct and incorrect responses were distinguished, $F(11, 3287) = 1.26$, $p = .24$, there was a significant interaction of correctness with situation, $F(11, 3287) = 2.59$, $p < .01$. Post-hoc tests clarified that significant differences emerged between incorrect and correct responses to two different paradigms in all but two cases: Only for “climb” and “shelf” did children who responded correctly to these situations responded faster than children responding wrong to the same situation.
Figure 10. Relationship between hit rates and mean RTs. Each data point shows the mean RT for one situation as a function of its hit rate. Means were calculated separately for the small- (circles) and the large picture-size subgroup (crosses) as well as for P3-help-side (blue), P2-help (green) and P1-bird (red). Dashed lines illustrate correlations between hit rates and RTs for the small-picture subgroup, solid lines for the large-picture subgroup. Correlation coefficients are shown adjacent to the corresponding regression lines. Asterisks mark significance according to Pearson’s correlation test: **p < .01, * p < .05. Regression lines were calculated by fitting linear models.
Data from small-picture subgroup

The detailed results of the ANOVAs for both, hit rates and RTs are given in Table 7. Hit rates as well as mean RTs for bird and human pictures are shown on the right side of Figure 9.

*Bird-depiction* was the only factor showing a significant main effect on hit rates in the small-picture subgroup, just as in the large-picture subgroup. In contrast to the large-picture subgroup, here, birds were more accurately recognized as such, MD = 4.56%, 95% CI[2.15%, 5.30%], p = .02. None of the other factors (*NoH-depiction, situation*) or any interaction influenced hit rates.

In this subgroup, I did not find a correlation between mean hit rates and RTs, r = .01, p = .98 (see red crosses and adjacent regression line in Fig. 10). This result still provides evidence that there is no speed-accuracy-trade-off in the small-picture subgroup, either. As in the large-picture subgroup, no effects of any factor (*NoH-depiction, bird-depiction, situation*) or their interactions were present for RTs.

What is more, I found the main effect of *correctness* to be significant in this subgroup, too, F(14, 1213) = 4.74, p = .03. Note that in contrast to the large-picture subgroup, not only the main effect of *situation* was non-significant, F(14, 1213) = 1.45, p = .12, but also its interaction with *correctness*, F(14, 1213) = 0.83, p = .64. This single difference in RTs is likely to be caused by the smaller sample size in the small-picture subgroup and does probably not reflect a systematic difference in effect patterns due to picture size.

Thus, the pattern of effects on RTs was the same in large- and small-picture subgroups despite one non-significant interaction. Regarding hit rates, the same factor *bird-depiction* yielded the only significant effect in both picture-size subgroups while no interactions were found in either picture-size subgroup. However, the main effect of *bird-depiction* had opposite
directions for large- and small-picture subgroups. Birds were categorized more accurately in the large-picture subgroup whereas humans were more accurately categorized in the small-picture subgroup (see Fig. 9). This inversion of bird-depiction’s main effect might arise due to differences in task difficulty according to picture-size: Hit rates were generally higher in the large-picture subgroup and on average above 92% (see section “Differences in response characteristics for the two picture-size subgroups”), indicating that a ceiling effect could have played a role in changing the direction of effect of bird-depiction on hit rates.

**P2-help: NoH- recognition after short stimulus presentation**

Data from large-picture subgroup

As opposed to P1-bird, in P2-help children had to decide whether they saw someone (a bird or a human) that needs help (NoH) or someone who does not need help (no-NoH). As in P1-bird, pictures were shown individually and for a duration of 500 ms. Thus, children had to base their decision on memory. The detailed results of the ANOVAs for both, hit rates and RTs, are given in Table 8. Figure 11 shows the mean hit rates and RTs for each situation.
Figure 11. Response characteristics per situation for the large-picture subgroup in P2-help. Hit rates (bars) and mean RTs (dots) were calculated across all variations of a given situation. Symbols indicate significant differences in hit rates (white symbols) and RTs (black symbols). The number of symbols corresponds to p-values of post-hoc Tukey HSD tests (one: p < .05, two: p < .01, three: p < .001). Asterisks mark significant differences compared to the situation “gap”, triangles to “table_chair” and crosses to “stair”. Error bars represent SEM. Note that differences are only indicated once for each pair of situations.

Table 8. Results of ANOVAs conducted in P2-help

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Large-Picture Subgroup</th>
<th></th>
<th>Small-Picture Subgroup</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hit Rates</td>
<td>RTs</td>
<td>Hit Rates</td>
<td>RTs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>p</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Situation</td>
<td>11</td>
<td>7.60***</td>
<td>&lt; .001</td>
<td>4.20***</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NoH-depiction</td>
<td>1</td>
<td>88.15***</td>
<td>&lt; .001</td>
<td>39.98***</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Bird-depiction</td>
<td>1</td>
<td>23.06***</td>
<td>&lt; .001</td>
<td>0.06</td>
<td>.81</td>
</tr>
<tr>
<td>Situation x NoH-depiction</td>
<td>11</td>
<td>3.36***</td>
<td>&lt; .001</td>
<td>1.24</td>
<td>.26</td>
</tr>
<tr>
<td>Situation x bird-depiction</td>
<td>11</td>
<td>1.99*</td>
<td>.03</td>
<td>0.62</td>
<td>.82</td>
</tr>
<tr>
<td>NoH-depiction x bird-depiction</td>
<td>1</td>
<td>75.84***</td>
<td>&lt; .001</td>
<td>0.13</td>
<td>.72</td>
</tr>
<tr>
<td>Situation x NoH-depiction x</td>
<td>11</td>
<td>1.60</td>
<td>.09</td>
<td>0.56</td>
<td>.86</td>
</tr>
<tr>
<td>Bird-depiction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ***p < .001; **p < .01
Hit rates in P2-help were profoundly influenced by the factors *situation*, *NoH-depiction* as well as *bird-depiction* and all two-way interactions. Note that this pattern clearly contrasts the results from P1-bird (same design, but different instruction: bird-human distinction required), where only *bird-depiction* influenced hit rates. Regarding differences between situations, post-hoc comparisons revealed that significant differences emerged only between “gap”, “table_chair”, “stair” and other situations (see white symbols Fig. 11). The results suggest that “gap” and “stair” can be considered as being more difficult to categorize in terms of NoH-recognition than other situations, even though the hit rates are still above chance. On the other hand, “table_chair” diverges from the other situations by means of being even more easily categorizable.

In general, NoH pictures were categorized better than no-NoH pictures, MD = 13.16%, 95% CI[15.90%, 10.41%], p < .001. Interestingly pictures of birds were also identified more accurately as showing NoH or no-NoH, MD = 6.83%, 95% CI[3.99%, 9.66%], compared to pictures of humans. The interaction of the factors *bird-* and *NoH-depiction* showed that the difference between the hit rates of bird and human depictions when the picture showed NoH was negligible, p = .09. The contingency table for this interaction is given in Table 9. In contrast, no-NoH pictures of birds were significantly better categorized than no-NoH-pictures of humans, p < .001. This also means that there were no differences in hit rates for pictures of humans depending on whether they showed someone in need for help or not, p = .07, while for pictures of birds this made a significant difference in a way that NoH depictions of birds were always correctly identified as such, whereas there was a considerable drop in response accuracy for no-NoH-depictions, p < .001.
Table 9. Mean hit rates and SDs illustrating the interaction of bird-depiction and NoH-depiction in the large-picture subgroup for P2-help.

<table>
<thead>
<tr>
<th></th>
<th>NoH-depiction</th>
<th>No-NoH-depiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-depiction</td>
<td>.79 (.41)</td>
<td>.74 (.44)</td>
</tr>
<tr>
<td>Bird-depiction</td>
<td>1.00 (.00)</td>
<td>.71 (.45)</td>
</tr>
</tbody>
</table>

*Note.* NoH = need-of-help. Numbers in brackets represent SDs.

In addition to the main effect of situation described above, showing that the situations “gap” and “stair” had lower hit rates compared to other situations, whereas “table_chair” had higher ones, an interaction of situation and NoH-depiction emerged, as illustrated in Figure 12. The NoH-depiction elicited significantly higher hit rates for the situations “blocks”, “boat”, “shelf” and “stair”. For the other nine situations no significant post-hoc differences for no-NoH/NoH-depictions regarding hit rate were observed.

The correlation between RTs and hit rates was significant and again negative, $r = -0.74$, $p < .01$, thus showing that there was no speed-accuracy-tradeoff (see green circles and adjacent regression line in Fig. 10), as demonstrated in earlier analyses, too. In contrast to hit rates, RTs were not affected by bird-depiction. On the other hand, main effects of situation as well as NoH-depiction were present for RTs, as well as for hit rates. In general, RTs were faster for NoH pictures, MD = 176 ms, 95% CI[121ms, 231 ms], $p < .001$. The four situations (“blocks”, “boat”, “shelf” and “stair”) for which post-hoc tests revealed significantly different RTs are marked with black symbols in Figure 11. All remaining main effects and interactions were non-significant.

As in P1-bird, RTs were significantly faster for correct compared to incorrect responses, $F(11, 3031) = 14.69$, $p < .001$. This finding indicates that the influence of correctness does not
depend on the content of the task given, but rather on the way in which the stimulus is presented and decision has to be made, as the same effect was also found in P1-bird. On the other hand, RTs also depended critically on situation in this paradigm, $F(11, 3031) = 4.19, p < .001$. The interaction of the factors situation and correctness was non-significant, $F(11, 3031) = 1.44, p = .15$.

![Graph](image)

**Figure 12.** Interaction effect of NoH-depiction and situation on hit rates. Data from the large-picture subgroup for paradigm P2-help is shown. This was the only paradigm and picture size subgroup for which the depicted interaction was significant. Filled black dots represent hit rates for NoH-depictions, hollow dots for no-NoH-depictions. Error bars represent SEM. Situations for which no-NoH/NoH-depictions were significantly different within the situation are marked with asterisks (*$p < .05$, ***$p < .001$, referring to post-hoc Tukey HSD tests). Also note that the main effect of NH-depiction nonetheless shows that NoH-depictions are generally recognized better.
Data from small-picture subgroup

The detailed results of the ANOVAs for both, hit rates and RTs are given in Table 8. Mean RTs and hit rates for the 14 situations remaining in analysis are shown in Figure 13.

Figure 13. Response characteristics per situation for the small-picture subgroup in P2-help. Hit rates (bars) and mean RTs (dots) were calculated across all variations of a given situation. White triangles mark significant differences in hit rates compared to the situation “table_chair”, their number corresponding to p-values of post-hoc Tukey HSD tests (one: p < .05, three: p < .001). Note that the ambiguous situation "gap" was excluded from analyses in this subgroup. Error bars represent SEM.

Hit rates again were profoundly influenced by situation, NoH-depiction and bird-depiction as well as the interaction of both latter factors, replicating the findings obtained in the large-picture subgroup. Post-hoc comparisons showed that significant differences between situations only emerged between the situation “table_chair” and others (see white triangles in Fig. 13).
These results underline the finding from the large-picture subgroup that “table_chair” seems to be more easily identifiable as depicting either NoH or no-NoH than the other situations, at least if the task requires to judge NoH out of memory after a brief stimulus presentation. Further replicating findings from the large-picture subgroup I again found responses to NoH-depictions to be more accurate, MD = 29.15%, 95% CI[33.90%, 24.40%], p < .001, as well as responses to bird-depictions, MD = 6.73%, 95% CI[1.82%, 11.63%], p < .01. It is noteworthy that all bird depictions that showed no-NoH were recognized with 100% accuracy (n = 181, see Table 10 for mean hit rates), just as in the large-picture subgroup. For human-depictions, however, hit rates were equal for no-NoH- and NoH-depictions. Instead of the interaction of bird-depiction and situation, I found the three-way interaction of all three factors to be significant. No further interaction reached significance.

Table 10. Mean hit rates and SDs illustrating the interaction of bird-depiction and NoH-depiction in the small-picture subgroup for P2-help.

<table>
<thead>
<tr>
<th></th>
<th>NoH-depiction</th>
<th>No-NoH-depiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-depiction</td>
<td>.83 (.38)</td>
<td>.61 (.49)</td>
</tr>
<tr>
<td>Bird-depiction</td>
<td>1.00 (.00)</td>
<td>.61 (.49)</td>
</tr>
</tbody>
</table>

*Note. NoH = need-of-help. Numbers in brackets represent SDs.*

Comparing RTs of more and less accurately categorized situations, I were again able to rule out a possible explanation of differences in either RTs or hit rates by speed-accuracy trade-off (see green crosses and adjacent regression line in Fig. 10). Pearson’s correlation was negative, r = -.55, p = .04, confirming previous observations. Also, as in the large-picture subgroup, the
same effect of NoH-depiction on RTs was found in this small-picture subgroup, showing that responses to NoH-pictures were made faster, MD = 153 ms, 95% CI[62 ms, 244 ms], p < .01. Again RTs were not influenced by either the factor bird-depiction, or the interaction of bird-depiction x NoH-depiction. There was no main effect of situation on RTs in this subgroup. Moreover analysis of data from the small picture subgroup also confirmed findings concerning differences in RTs regarding correct and incorrect responses: Children were faster when giving correct compared to incorrect responses, F(13, 1041) = 6.86, p < .01. There was no main effect of situation, F(13, 1041) = 1.14, p = .32, and also no interaction of situation and correctness, F(13, 1041) = 1.15, p = .31.

As in P1-bird I did not find pronounced differences in main effects between small- and large-picture subgroups in P2-help. Divergences in effect patterns regarding hit rates were restricted to the absence of some significant interactions in the small picture size subgroup, which was also considerably smaller in sample size (N = 22) than the large-picture subgroup (N = 60). Regarding RTs, the main effect of situation did not emerge in the smaller small-picture subgroup. For hit rates and RTs, the absence of interactions and of one main effect in the small-compared to the large-picture subgroup are likely due to the smaller sample size, since all effects for both pictures size-subgroups pointed in the same direction.

**P3-help-side: direct no-NoH / NoH comparison without time restriction**

In P3-help-side, children had to distinguish between NoH and no-NoH, as in P2-help, but picture presentation and decision mode differed considerably: Whereas individual pictures were shown only briefly (500 ms) in P2-help, corresponding no-NoH/NoH-pictures-pairs were presented side by side in P3-help until the child made a response without any time restriction of presentation or response.
Data from larger picture presentation

The detailed results of the ANOVAs for both, hit rates and RTs are given in Table 11. Mean hit rates and RTs for the different situations are presented in Figure 14.

**Figure 14.** Response characteristics per situation for the large-picture subgroup in P3-help-side. Hit rates (bars) and mean RTs (dots) were calculated across all variations of a given situation. Symbols indicate significant differences in hit rates (white symbols) and RTs (black symbols). The number of symbols corresponds to p-values of post-hoc Tukey HSD tests (one: \( p < .05 \), two: \( p < .01 \), three: \( p < .001 \)). Asterisks mark significant differences compared to the situation “gap”, triangles to “table_chair” and diamonds to “shirt”. Error bars represent SEM. Note that differences are only indicated once for each pair of situations.
Table 11. Results of ANOVAs conducted in P3-help-side

<table>
<thead>
<tr>
<th>Source</th>
<th>Large-Picture Subgroup</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Small-Picture Subgroup</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>F</td>
<td>p</td>
<td>F</td>
<td>p</td>
<td>df</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Situation</td>
<td>10</td>
<td>4.67***</td>
<td>&lt; .001</td>
<td>6.93</td>
<td>&lt; .001</td>
<td>13</td>
<td>2.09*</td>
<td>.01</td>
</tr>
<tr>
<td>NH-side</td>
<td>1</td>
<td>4.54*</td>
<td>.03</td>
<td>1.33</td>
<td>.25</td>
<td>1</td>
<td>0.40</td>
<td>.53</td>
</tr>
<tr>
<td>Bird-depiction</td>
<td>1</td>
<td>0.85</td>
<td>.36</td>
<td>4.24*</td>
<td>.04</td>
<td>1</td>
<td>0.35</td>
<td>.56</td>
</tr>
<tr>
<td>Situation x NH-side</td>
<td>10</td>
<td>0.92</td>
<td>.52</td>
<td>1.20</td>
<td>.28</td>
<td>13</td>
<td>0.40</td>
<td>.97</td>
</tr>
<tr>
<td>Situation x bird-depiction</td>
<td>10</td>
<td>0.78</td>
<td>.64</td>
<td>1.55</td>
<td>.12</td>
<td>1</td>
<td>0.53</td>
<td>.91</td>
</tr>
<tr>
<td>NH-side x bird-depiction</td>
<td>1</td>
<td>6.35*</td>
<td>.01</td>
<td>0.40</td>
<td>.53</td>
<td>1</td>
<td>0.23</td>
<td>.63</td>
</tr>
<tr>
<td>Situation x NH-side x bird-depiction</td>
<td>10</td>
<td>0.93</td>
<td>.51</td>
<td>1.18</td>
<td>.30</td>
<td>13</td>
<td>0.96</td>
<td>.49</td>
</tr>
</tbody>
</table>

*Note.***p < .001; *p < .05
Situation was one of two factors that had a main effect on hit rates. Post-hoc Tukey HSD tests showed that accuracy differences were only significant between the situation “gap” and most other situations (see white asterisks Fig. 14). This led to the assumption that the situation “gap” is systematically harder to differentiate than some easiest to differentiate situations, even though children were generally able to tell whether pictures of the situation “gap” depicted NoH or no-NoH (see section “Ambiguous situations”).

A main effect of NoH-side demonstrated that NoH-depictions were more accurately categorized as such, when shown on the right side of the screen. However, an interaction of bird-depiction and NoH-side relativized this finding: Post-hoc TukeyHSD revealed that significant differences in response accuracy for picture-pairs in which NoH-pictures shown on the left vs. right side emerged only for pictures of birds, not for those of humans (see Fig. 15 for an illustration of this interaction). This means, that for NoH-pictures of humans the position on the screen did not influence hit rates, while NoH-pictures of birds were better recognized when shown on the right side of the screen. All remaining interactions of the ANOVA were non-significant.

A test of Pearson's correlation coefficient confirmed again that there was no speed-accuracy-trade-off and revealed on the contrary that in general RTs were shorter for situations that were categorized more accurately, \( r = -.76, p < .01 \) (see blue circles and adjacent regression line in Fig. 10).
Figure 15. Interaction of side of need-of-help presentation (NoH-side) and bird-depiction. Hit rates are shown across situations for responses in the large-picture subgroup in P3-help-side. Hollow circles connected with a dashed line represent data for human-depictions, filled circles connected with a solid line for bird-depictions. Error bars represent SEM. Note that the only significant difference in post-hoc tests is the one between NoH-depictions of birds shown on the left vs. right side, p < .01, all remaining p ≥ .07.

The ANOVA conducted on RTs yielded a different pattern of results compared to the one for hit rates. NoH-side had no main effect on RTs. In contrast, not only situation but also bird-depiction did influence mean RTs independently from each other: RTs were slower for pictures of birds, MD = 70 ms, 95% CI[2 ms, 137 ms], p = .04. Post-hoc comparison of RTs for different situations did not correspond to comparisons of hit rates for different situations: Mean RT for the “gap” situation was sometimes equal and sometimes slower compared to situations that were significantly better categorized than “gap” (compare presence of white and black asterisks in Fig. 14). What is more, there were also significant differences in RTs between other situations that did not differ regarding hit rates, i.e. “table-chair” and “shirt” (see black triangles and
diamonds, Fig. 14). Both findings indicate that RTs are sensitive to some other aspect of help recognition than hit rates. Correctness did not influence RTs in this paradigm, in which there were no time constraints for stimulus presentation, F(10, 1654) = 0.03, p = .85, as opposed to the other two time-constrained paradigms. In contrast, the main effect of situation was significant, F(10, 1654)= 6.88, p < .001 (see Fig. 14). The interaction of situation and correctness was non-significant, F(10, 1654) = 1.16, p = .31.

Data from smaller picture presentation

The detailed results of the ANOVAs for both, hit rates and RTs are given in Table 11. Mean hit rates and RTs are presented in Figure 16.

For the small-picture subgroup, the only significant effect was the one of situation. However, for this smaller (N = 22) subgroup, Tukey HSD tests could not reveal any significantly different pairs of situations. Nonetheless, the pattern of mean hit rates resembles the one obtained in the bigger large-picture subgroup (compare bars of Fig. 14 and Fig. 16). As for the large-picture subgroup, no speed-accuracy trade-off was found, as proved by a significant negative correlation between mean RTs and hit rates of the different situations, r = -.56, p = .05 (see blue crosses and adjacent regression line in Fig. 10).

The pattern of effects on RTs was the same in the small- compared to the large-picture subgroup. RTs were most strongly influenced by situation and again, differences emerged between “table_chair” and other situations (see asterisks Fig. 16). As for the large-picture subgroup, RTs were significantly slower (p = .04) to bird pictures, M = 1802 ms, SD = 776 ms, compared to pictures of humans, M = 1701 ms, SD = 753 ms. All remaining effects were non-significant. Also in line with the results obtained in the large-picture subgroup, RTs were not
influenced by correctness of the children's responses, main effect: F(12, 591) = 0.32, p = .57; interaction: F(12, 591) = 1.05, p = .40.

Effect patterns found in the large and small picture-size subgroups in paradigm P3-help were highly similar. Given that considerably fewer children (N = 22) were assigned to the small picture-size subgroup and that the only differences were restricted to a lack of significant main effects and interactions in this subgroup, it is likely that content-related picture properties' influences on response characteristics in P3-help-side were independent of picture size. What is more, I also found that differences between situations were similar in both subgroups, the only discrepancy being again that some differences did not reach significance in the smaller subgroup.

Figure 16. Response characteristics per situation for the small-picture subgroup in P3-help-side. Hit rates (bars) and mean RTs (dots) were calculated across all variations of a given situation. Symbols indicate significant differences in RTs between the situation “table_chair” and others. The number of symbols corresponds to p-values of post-hoc Tukey HSD tests (one: p < .05, two: p < .01). Error bars represent SEM.
Changes in hit rates and RTs according to paradigm sequence

As it is plausible that prior exposure to pictures of the stimulus set affects response characteristics and that this effect can be modified by the task demands during prior exposure, I assessed whether response characteristics were generally influenced by the sequence in which children absolved paradigms. Even though included as factor in the ANOVAs, I will not report main effects of paradigm separately, here, as they followed the same pattern as already reported above.

Data from large-picture subgroup

After excluding data of ambiguous situations and of unplanned paradigm sequences, 3143, 2731 and 1558 trials remained for analyses in P1-bird, P2-help and P3-help-side respectively. Detailed results of all ANOVAs conducted for this subgroup are shown on the left sides of Tables 12-14. Effects of paradigm sequence on hit rates and RTs of the large-picture subgroup are illustrated in the top row of Figure 17.

When pictures were presented in larger size, hit rates were higher for all paradigms if P1-bird had been absolved first. RTs in P1-bird and P2-help RTs were lower when P1-bird had been absolved first. No significant difference between RTs according to paradigm sequence was found for P3-help-side. All interactions were non-significant.
Figure 17. Hit rates and mean RTs according to paradigm sequence for each paradigm. Red bars represent data of children who did P1-bird first, green bars of children who did P2-help first. Data for the large-picture subgroup is shown on top (A, B), data for the small-picture subgroup is shown on the bottom (C, D). Hit rates are shown in the left graphs (A, C), RTs in the right graphs (B, D). Asterisks mark significant differences as indicated by post-hoc Tukey HSD tests: ***p < .001; **p < .01; *p < .05. Note that significant main effects of paradigm sequence point into the same direction in both picture-size subgroups for hit rates and RTs.
Table 12. Results of ANOVAs investigating the influence of *paradigm sequence* on hit rates and RTs in P3-help-side

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>10</td>
<td>3.89***</td>
<td>&lt; .001</td>
<td>6.44***</td>
<td>&lt; .001</td>
<td>13</td>
<td>2.50**</td>
<td>&lt; .01</td>
<td>4.54***</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Paradigm sequence</td>
<td>1</td>
<td>8.36**</td>
<td>&lt; .01</td>
<td>5.62*</td>
<td>.02</td>
<td>1</td>
<td>1.25</td>
<td>.26</td>
<td>7.33**</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Situation x Paradigm sequence</td>
<td>10</td>
<td>0.98</td>
<td>.46</td>
<td>0.65</td>
<td>.77</td>
<td>13</td>
<td>2.72***</td>
<td>&lt; .001</td>
<td>1.56</td>
<td>.09</td>
</tr>
</tbody>
</table>

*Note.* *** p < .001; **p < .01; *p < .05

Table 13. Results of ANOVAs investigating the influence of *paradigm sequence* on hit rates and RTs in P2-help

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>11</td>
<td>6.01***</td>
<td>&lt; .001</td>
<td>4.48***</td>
<td>&lt; .001</td>
<td>13</td>
<td>1.75*</td>
<td>.05</td>
<td>1.60</td>
<td>.08</td>
</tr>
<tr>
<td>Paradigm sequence</td>
<td>1</td>
<td>9.99**</td>
<td>&lt; .01</td>
<td>10.06**</td>
<td>&lt; .01</td>
<td>1</td>
<td>1.12</td>
<td>.29</td>
<td>7.13**</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Situation x Paradigm sequence</td>
<td>11</td>
<td>0.49</td>
<td>.91</td>
<td>0.62</td>
<td>.81</td>
<td>13</td>
<td>0.85</td>
<td>.60</td>
<td>0.95</td>
<td>.50</td>
</tr>
</tbody>
</table>

*Note.* *** p < .001; **p < .01; *p < .05

Table 14. Results of ANOVAs investigating the influence of *paradigm sequence* on hit rates and RTs in P1-help

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-Picture Subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hit Rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>11</td>
<td>0.94</td>
<td>.50</td>
<td>1.16</td>
<td>.31</td>
<td>14</td>
<td>0.20</td>
<td>1.00</td>
<td>1.30</td>
<td>.20</td>
</tr>
<tr>
<td>Paradigm sequence</td>
<td>1</td>
<td>7.70**</td>
<td>&lt; .01</td>
<td>0.10</td>
<td>.75</td>
<td>1</td>
<td>1.80</td>
<td>.18</td>
<td>20.24***</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Situation x Paradigm sequence</td>
<td>11</td>
<td>1.53</td>
<td>.11</td>
<td>1.33</td>
<td>.20</td>
<td>14</td>
<td>0.50</td>
<td>.93</td>
<td>0.27</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note.* *** p < .001; **p < .01
**Data from small-picture subgroup**

After excluding data of ambiguous situations and of unplanned paradigm sequences, 961, 886 and 556 trials remained for analyses in P1-bird, P2-help and P3-help side respectively. Detailed results of all ANOVAs conducted for this subgroup are shown on the right sides of Tables 12-14. Effects of *paradigm sequence* on hit rates and RTs for the small-picture subgroup are illustrated in the bottom row of Figure 17.

As opposed to the large-picture subgroup, *paradigm sequence* had no main effect on hit rates in the small-picture subgroup. In contrast, the same main effect of *paradigm sequence* on RTs as found for P1-bird and P2-help in the large-picture subgroup emerged for all paradigms in this subgroup. RTs were generally lower if P1-bird had been done first. There was also one significant interaction of *situation* and *paradigm sequence* in the small-picture subgroup, affecting hit rates in P3-help-side: Only for one situation, i.e. “blocks”, hit rates were significantly higher if P2-help was done first, all other significant post-hoc differences concerned diverging hit rates between situations.

**Summary**

In sum, *paradigm sequence* had consistent effects on response characteristics across picture-size subgroups. Hit rates were significantly higher if a human-bird-distinction was required first and if stimuli were presented in larger size. Consistent with the negative correlation between hit rates and RTs I found in stimulus content analyses, RTs were lower if P1-bird was done first in all but one case (P3-help-side in large-picture presentation). *Paradigm sequence* interacted with a picture content related factor only once, i.e. regarding hit rates in P3-help in the small-picture size subgroup. Thus, the main effect of *paradigm sequence* can primarily be
regarded as a superimposed factor influencing response characteristics, not influencing analyses of picture content.
Discussion

This thesis introduces the NeoHelp: a new set of standardized visual stimuli that can be employed for studying need-of-help recognition as an aspect of socioemotional development and a precondition of active helping behavior. The main purpose of this thesis was to test empirically whether the NeoHelp stimulus set meets the criteria similarity, clarity and recognizability considered during its creation. The picture pairs that comprise the stimulus set are perceptually highly similar and yet unambiguous concerning the detection of either need-of-help (NoH) or no-need-of-help (no-NoH) in all but two cases. To ensure that developmental aspects of need-of-help recognition can also be investigated using NeoHelp, the stimuli are comprehensive and interesting for children. They can be used in a variety of experimental designs, as has been exemplified with three different paradigms. Data of response characteristics aiding stimulus selection for future studies and allowing direct comparison of results is provided by means of hit rates and mean RTs from children in a wide age range (three to 13 years, see appendix for a summary of the complete data). Overall, results a) demonstrate that the pictures’ content is correctly distinguished, b) prove that different tasks can be successfully applied using the NeoHelp stimulus set and c) provide a standard of response characteristics for selection of stimuli and comparison in future studies. Thus, this thesis shows that the NeoHelp is suitable for behavioral testing. Moreover it demonstrates that the picture pairs comprising the NeoHelp are highly similar regarding low-level picture properties, rendering them suitable for psychophysiological testing as well. The NeoHelp thus overcomes the limitation of previously used visual stimulus material depicting social situations by providing control of lower level picture properties and data concerning recognisability of content as well as exemplary response characteristics of a normative population.
Currently, the NeoHelp-stimulus set consists of pictures for 15 different situations and their variations, a total of 82 individual stimuli (see appendix for a complete list). There is evidence that all picture pairs are characterized by a high structural similarity (SSIM) and thus high similarity of perceptual properties (see section “Picture similarity” of the results as well as appendix for detailed values). Thus, the two content categories “need-of-help” and “no-need-of-help” as well as the content categories “human” and “bird” are perceptually very similar. Therefore, the NeoHelp pictures can be used without the risk of built-in (low level) perceptual differences leading to differences in response characteristics that might be confused with need-of-help content effects or effects of either human or bird depiction.

Despite the high similarity on a lower perceptual level, children recognized whether need-of-help or no-need-of-help is depicted for most situations. Only two situations, “climb” and “gap”, turned out to be ambiguous concerning need-of-help-content under certain conditions (see Fig. 8). Slight variations in human depictions of a given situation (i.e. of gender or age or race of the child depicted) did not alter response characteristics, thus providing the opportunity to increase the number of trials per situation without repeatedly showing the same stimulus. As a control condition the NeoHelp stimulus set includes depictions of birds in situations analog to the human NoH-no-NoH-picture pairs. The depiction of humans vs. birds was unambiguous for all situations, children recognized whether a bird or a human was depicted highly reliably even when pictures are presented for only 500 ms and a decision was taken after stimulus offset (see Fig. 8 C). In light of these results, it can be concluded that the NeoHelp stimulus set is well suited not only for any kind of behavioral testing, but also for the use in psychophysiological studies, such as EEG, eye tracking or fMRI.
This thesis gives an example of three different tasks with varying demands (see Fig. 3 for an overview of the paradigm sequences). The three paradigms were created in a way that allows disentangling the influences of need-of-help recognition abilities from those of the stimulus presentation mode. The results demonstrate that both, restricted stimulus presentation and need-of-help recognition lead to characteristic changes in hit rates as well as RTs. In the paradigm with the highest task load (P2-help: need-of-help recognition and time restriction and decision from memory), hit rates were lowest and RTs were longest. Changing the task to an (easier) human vs. bird distinction or allowing open-ended pair-wise picture comparison both shortened RTs and increased hit rates, indicating that these effects are generally caused by lowering task demands. It is noteworthy that for hit rates, changing task content had a greater effect than prolonging stimulus presentation, whereas RTs depended more strongly on the mode of stimulus presentation (see Fig. 6 for a summary). These results provide evidence that the NeoHelp stimuli elicit coherent and predictable changes in response characteristics depending on task demands. As a consequence, a wide range of experimental designs can be applied to the NeoHelp stimulus set, allowing extraction of effects of need-of-help recognition abilities. In the example paradigms used here, this can be achieved by considering effects emerging in both help-content related paradigms (P-2-help and P3-help-side), but not in P1-bird.

Distinct task demands also lead to discernible effect patterns regarding influences of picture properties, as illustrated in Figure 18. Response characteristics did not differ across situations when a human-bird distinction was required (see Table 7), but did so systematically in both NoH-related paradigms (see Tables 8 and 9). Thus, the NeoHelp provides stimuli that differ in their difficulty regarding need-of-help recognition while there is no indication that situations that differ in their need-of-help-related clarity of content differ in their overall comprehensibility.
This offers the opportunity to select specific situations for testing, depending on the expected ability of the study population to recognize need-of-help.

*Figure 18.* Overview of main effects on hit rates and mean RTs per paradigm. Factors that had a significant main effect are written inside ellipses. Red shaded ellipses indicate relationships with the task demand “need-of-help-recognition”, green shaded areas indicate relationships with the task demand “short picture presentation and decision from memory” and overlapping shading indicate that both task demands had an influence. White ellipses illustrate that the effect was strictly paradigm-specific. Ellipses located at the top refer to main effects on hit rates only, ellipses located at the bottom to main effects on RTs only. Ellipses that lie on the dashed line separating top and bottom panel show factors that exerted main effects on hit rates as well as RTs.
At this point, however, which characteristics of a situation make it more or less easy for children to categorize can only be speculated. It is likely that a higher amount of experience with certain situations increased the ability of children to detect subtle changes illustrating the need of help [47,48]. On the other hand it is also possible that some aspect of the illustration of the situation might have influenced its clarity. To be able to draw definite conclusions about the reasons for differences between situations, the NeoHelp stimulus set needs to be extended and supplemented with ratings of familiarity and/or frequency of occurrence. Systematic analyses of population characteristics and their relationship with success of categorization will also provide meaningful insights.

The need-of-help recognition process was also malleable to changes in picture-size, albeit differences were only evident in RTs, which were shorter for smaller pictures. This difference is most likely attributable to a higher rate of fixations in larger pictures, resulting in a higher amount of information to be integrated in this condition [49]. The effect is not present for the human-bird distinction task probably due to the fact that there, categorization is easier because the boundaries between the categories human and bird are sharper and the distinction between these categories is learned early in development [50,51]. Whether picture size influenced hit rates on the other hand depended on the type of stimulus presentation: Hit rates were higher for larger pictures only if stimulus presentation was restricted to 500 ms and one picture.

One more finding, which should be considered when using the NeoHelp stimulus set, is that the sequence in which different tasks are given for the same pictures influences response characteristics (see Fig. 17). Several reasons for this finding have to be taken into consideration: First, at least paradigms with short stimulus presentation are malleable to influences of prior exposure to the stimulus material, as indicated by differences in hit rates and RTs in P1-bird and
P2-help according to *paradigm sequence*. Second, even if the amount of prior exposure is held constant, as was the case for P3-help-side, the sequence in which other tasks have been assigned to the same stimuli affects response characteristics in the following task. One plausible explanation for the effects is that children were less motivated when the task demands were too high in the beginning, as might be the case for P2-help. To draw more definite conclusions about the effects of prior stimulus exposure and its relationship with task demands, one would need a specifically designed experiment.

To sum up, our findings confirm that children are able to recognize need-of-help as well as the depiction of a human vs. a bird reliably on most pictures of the newly created stimulus set NeoHelp. When children were asked to make different judgments about specific aspects of the identical pictures, different response patterns were obtained. The task given determined which picture properties influenced response characteristics. These task-specific effects of picture content were insensitive to additional variations of gender, age or ethnicity in human depictions. An additional factor influencing response characteristics was the sequence in which tasks were absolved. Another interesting finding was that the obtained effect patterns were consistently different for RTs and hit rates. It remains for future studies, however, to illuminate which aspects of help-recognition are mirrored in hit rates and RTs.

One major limitation of the results presented here is that they do not take into consideration child-related variables such as age, gender or pet-ownership which were also collected. Also the relation between data obtained about need-of-help recognition gained with the NeoHelp and different social experiences and other environmental factors, personality variables and clinical symptoms has to be examined. Moreover an external validation of the stimulus set, taking into consideration children’s willingness to help and their actual helping behavior, is needed. As it is
the purpose of this thesis to introduce the NeoHelp stimulus set as a research tool, these issues will have to be addressed in future studies. One great advantage for future studies with the NeoHelp stimulus set is neurological and psychophysiological mechanisms underlying NoH-recognition can also be assessed. After such further development, the NeoHelp stimulus set might also be useful as a diagnostic indication within the clinical assessment of socioemotional impairments, such as autism.

In conclusion this thesis represents the first step for establishing the NeoHelp stimulus set as a new research tool to assess one specific aspect of socioemotional development, i.e. need-of-help recognition. The process of developing the NeoHelp stimulus set, however, is far from complete. By employing the same stimulus creation procedure described in this thesis (see “Stimuli” in the methods section), the NeoHelp can be expanded in any direction by creating more and more diverse situations, variations of human depictions and it is highly probable that this is also true for new kinds of control pictures (e.g. other non-human beings, greebles, objects). Pictures are available in formats for direct use as stimuli as well as for their modification into new variations. I therefore encourage all readers to contact me with their suggestions or if they wish to work with the stimulus files.
References


Picture similarity, hit rates and mean RTs for individual pictures of the NeoHelp Stimulus Set.
All individual pictures will be listed in alphabetical order of the situations’ names in the following structure:

“Situation X”

First picture pair of situation X

Specification of picture properties [human/bird; further specification of variation in gender, age, etc.]

NoH-distinction [data for a time restricted need-of-help-distinction task]
  n = data sets available for the picture considered
  hit rate = proportion of correct responses (SD)
  RT = mean response time in ms (SD)

human-bird-distinction [data for a time restricted human-bird-distinction task]
  [...]

NoH-distinction (pairwise comparison)
  n =
  hit rate = . ()
  RT = ()

human-bird-distinction
  n =
  hit rate = . ()
  RT = ()

SSIM [the index used for picture similarity]
  reference picture = 0.
  NoH-depiction = 0.

NoH-distinction (pairwise comparison)
  [data for a need-of-help-distinction task with pairwise picture comparison and unlimited picture presentation]
  [...]

variation of situation X

new variation of situation X, not empirically tested yet

“Situation Y”
NoH-distinction
n = 12
hit rate = .92 (.29)
RT = 940 (401)

human-bird-distinction
n = 17
hit rate = .88 (.33)
RT = 772 (547)

NoH-distinction (pairwise comparison)
n = 16
hit rate = .75 (.45)
RT = 1776 (582)

SSIM
reference picture = 0.89
NoH-depiction = 0.91

NoH-distinction
n = 13
hit rate = .38 (.51)
RT = 1497 (659)

human-bird-distinction
n = 17
hit rate = .88 (.33)
RT = 726 (451)

human; school age; girl

NoH-distinction
n = 15
hit rate = 1.00 (0.00)
RT = 661 (471)

human-bird-distinction
n = 13
hit rate = .92 (.28)
RT = 661 (471)

NoH-distinction (pairwise comparison)
n = 16
hit rate = .94 (.25)
RT = 2249 (895)

SSIM
reference picture = 1.00
NoH-depiction = 0.88

bird

NoH-distinction
n = 13
hit rate = 1.00 (0.00)
RT = 1036 (619)

human-bird-distinction
n = 15
hit rate = .93 (.26)
RT = 737 (499)

NoH-distinction (pairwise comparison)
n = 16
hit rate = .93 (.26)
RT = 2048 (631)

SSIM
reference picture = 0.89
NoH-depiction = 0.91

NoH-distinction
n = 15
hit rate = .40 (.51)
RT = 1382 (871)

human-bird-distinction
n = 13
hit rate = .92 (.28)
RT = 830 (490)

human; school age; boy

NoH-distinction
n = 15
hit rate = .63 (.50)
RT = 1447 (773)

human-bird-distinction
n = 13
hit rate = .92 (.28)
RT = 830 (490)

NoH-distinction (pairwise comparison)
n = 16
hit rate = .75 (.45)
RT = 1776 (582)

SSIM
reference picture = 0.89
NoH-depiction = 0.91
“blocks”

human; kindergarden age; girl

NoH-distinction
\[ n = 67 \]
\[ \text{hit rate} = .76 (.43) \]
\[ \text{RT} = 1226 (721) \]

human-bird-distinction
\[ n = 70 \]
\[ \text{hit rate} = .96 (.20) \]
\[ \text{RT} = 786 (445) \]

NoH-distinction (pairwise comparison)
\[ n = 74 \]
\[ \text{hit rate} = .88 (.33) \]
\[ \text{RT} = 2248 (895) \]

SSIM
reference picture = 1.00
NoH-depiction = 0.81

human; kindergarden age; boy

NoH-distinction
\[ n = 63 \]
\[ \text{hit rate} = .86 (.35) \]
\[ \text{RT} = 1175 (648) \]

human-bird-distinction
\[ n = 67 \]
\[ \text{hit rate} = .93 (.26) \]
\[ \text{RT} = 816 (453) \]

NoH-distinction (pairwise comparison)
\[ n = 72 \]
\[ \text{hit rate} = .90 (.30) \]
\[ \text{RT} = 1784 (582) \]

SSIM
reference picture = 0.81
NoH-depiction = 0.81

bird

NoH-distinction
\[ n = 59 \]
\[ \text{hit rate} = 1.00 (0.00) \]
\[ \text{RT} = 1234 (770) \]

human-bird-distinction
\[ n = 67 \]
\[ \text{hit rate} = .88 (.33) \]
\[ \text{RT} = 801 (429) \]

NoH-distinction (pairwise comparison)
\[ n = 72 \]
\[ \text{hit rate} = .90 (.30) \]
\[ \text{RT} = 1842 (631) \]

SSIM
reference picture = 0.82
NoH-depiction = 0.86
“boat”

human; kindergarten age; boy

NoH-distinction
n = 64
hit rate = .81 (.39)
RT = 1303 (703)

human-bird-distinction
n = 67
hit rate = .93 (.26)
RT = 794 (484)

NoH-distinction (pairwise comparison)
n = 75
hit rate = .91 (.29)
RT = 1817 (727)

SSIM
reference picture = 1.00
NoH-depiction = 0.76

human; kindergarten age; girl

NoH-distinction
n = 64
hit rate = .80 (.41)
RT = 1240 (772)

human-bird-distinction
n = 69
hit rate = .94 (.24)
RT = 741 (478)

NoH-distinction (pairwise comparison)
n = 73
hit rate = .92 (.30)
RT = 1674 (598)

SSIM
reference picture = 0.85
NoH-depiction = 0.74

bird

NoH-distinction
n = 53
hit rate = 1.00 (0.00)
RT = 1248 (794)

human-bird-distinction
n = 72
hit rate = .94 (.23)
RT = 775 (449)

NoH-distinction (pairwise comparison)
n = 72
hit rate = .92 (.28)
RT = 1842 (756)

SSIM
reference picture = 0.86
NoH-depiction = 0.89
“branch”

human; school age; boy

NoH-distinction
n = 67
hit rate = .82 (.39)
RT = 1086 (749)

human-bird-distinction
n = 68
hit rate = .93 (.26)
RT = 797 (481)

NoH-distinction (pairwise comparison)
n = 72
hit rate = .97 (.17)
RT = 1530 (582)

SSIM
reference picture = 1.00
NoH-depiction = 0.87

human; school age; girl

NoH-distinction
n = 63
hit rate = .86 (.35)
RT = 1078 (644)

human-bird-distinction
n = 69
hit rate = .96 (.21)
RT = 790 (514)

NoH-distinction (pairwise comparison)
n = 79
hit rate = .96 (.19)
RT = 1716 (638)

SSIM
reference picture = 0.91
NoH-depiction = 0.87

bird

NoH-distinction
n = 59
hit rate = 1.00 (0.00)
RT = 910 (527)

human-bird-distinction
n = 70
hit rate = .90 (.30)
RT = 796 (480)

NoH-distinction (pairwise comparison)
n = 71
hit rate = .92 (.28)
RT = 1721 (648)

SSIM
reference picture = 0.89
NoH-depiction = 0.90
“bucket” (data for small picture presentation only)

human; school age; boy

- NoH-distinction
  - n = 14
  - hit rate = .71 (.47)
  - RT = 1092 (834)

- human-bird-distinction
  - n = 13
  - hit rate = .92 (.28)
  - RT = 1025 (766)

- NoH-distinction (pairwise comparison)
  - n = 14
  - hit rate = .71 (.47)
  - RT = 1880 (1022)

SSIM

- reference picture = 1.00
- NoH-depiction = 0.90

human; school age; girl

- NoH-distinction
  - n = 15
  - hit rate = .73 (.46)
  - RT = 1338 (931)

- human-bird-distinction
  - n = 15
  - hit rate = .80 (.41)
  - RT = 816 (546)

- NoH-distinction (pairwise comparison)
  - n = 12
  - hit rate = .83 (.39)
  - RT = 1948 (1121)

SSIM

- reference picture = 0.90
- NoH-depiction = 0.89

bird

- NoH-distinction
  - n =
  - hit rate = .
  - RT = 

- human-bird-distinction
  - n =
  - hit rate = .
  - RT = 

- NoH-distinction (pairwise comparison)
  - n = 14
  - hit rate = .79 (.43)
  - RT = 2308 (1071)

SSIM

- reference picture = 0.88
- NoH-depiction = 0.89
“climb”

human; toddler

<table>
<thead>
<tr>
<th>Task</th>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>Hit rate</td>
<td>.65 (.48)</td>
<td>.93 (.26)</td>
</tr>
<tr>
<td>RT</td>
<td>1369 (750)</td>
<td>885 (442)</td>
</tr>
</tbody>
</table>

NoH-distinction (pairwise comparison)

<table>
<thead>
<tr>
<th>n</th>
<th>Hit rate</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>.52 (.33)</td>
<td>2106 (958)</td>
</tr>
</tbody>
</table>

SSIM

<table>
<thead>
<tr>
<th>SSIM</th>
<th>reference picture</th>
<th>NoH-depiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>0.91</td>
<td>reference picture</td>
<td>NoH-depiction</td>
</tr>
</tbody>
</table>

bird

<table>
<thead>
<tr>
<th>Task</th>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>42</td>
<td>70</td>
</tr>
<tr>
<td>Hit rate</td>
<td>1.00 (0.00)</td>
<td>.89 (.32)</td>
</tr>
<tr>
<td>RT</td>
<td>1374 (859)</td>
<td>797 (532)</td>
</tr>
</tbody>
</table>

NoH-distinction (pairwise comparison)

<table>
<thead>
<tr>
<th>n</th>
<th>Hit rate</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>.62 (.49)</td>
<td>1897 (862)</td>
</tr>
</tbody>
</table>

SSIM

<table>
<thead>
<tr>
<th>SSIM</th>
<th>reference picture</th>
<th>NoH-depiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>reference picture</td>
<td>NoH-depiction</td>
</tr>
<tr>
<td>0.88</td>
<td>reference picture</td>
<td>NoH-depiction</td>
</tr>
</tbody>
</table>
“door”

human; school age; boy

<table>
<thead>
<tr>
<th>NoH-distinction (pairwise comparison)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 72</td>
<td>reference picture = 1.00</td>
</tr>
<tr>
<td>hit rate = .96 (.20)</td>
<td>NoH-depiction = 0.89</td>
</tr>
<tr>
<td>RT = 1777 (672)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 69</td>
<td>n = 62</td>
</tr>
<tr>
<td>hit rate = .87 (.34)</td>
<td>hit rate = .95 (.22)</td>
</tr>
<tr>
<td>RT = 1013 (694)</td>
<td>RT = 732 (358)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 66</td>
<td>n = 62</td>
</tr>
<tr>
<td>hit rate = .73 (.45)</td>
<td>hit rate = .95 (.22)</td>
</tr>
<tr>
<td>RT = 1291 (652)</td>
<td>RT = 798 (441)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction (pairwise comparison)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 71</td>
</tr>
<tr>
<td>hit rate = .93 (.26)</td>
</tr>
<tr>
<td>RT = 1693 (574)</td>
</tr>
</tbody>
</table>

bird

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 59</td>
<td>n = 64</td>
</tr>
<tr>
<td>hit rate = 1.00 (0.00)</td>
<td>hit rate = .88 (.33)</td>
</tr>
<tr>
<td>RT = 1142 (773)</td>
<td>RT = 667 (406)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 66</td>
<td>n = 70</td>
</tr>
<tr>
<td>hit rate = .79 (.41)</td>
<td>hit rate = .86 (.35)</td>
</tr>
<tr>
<td>RT = 1242 (697)</td>
<td>RT = 782 (511)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction (pairwise comparison)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 71</td>
</tr>
<tr>
<td>hit rate = .93 (.26)</td>
</tr>
<tr>
<td>RT = 1693 (574)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference picture = 0.90</td>
</tr>
<tr>
<td>NoH-depiction = 0.83</td>
</tr>
</tbody>
</table>
human; school age; girl

NoH-distinction
no data available yet

human-bird-distinction
no data available yet

NoH-distinction (pairwise comparison)
no data available yet

SSIM
reference picture =
NoH-depiction =
NoH-distinction
n = 15
hit rate = .87 (.35)
RT = 893 (499)

human-bird-distinction
n = 15
hit rate = .67 (.48)
RT = 926 (661)

NoH-distinction (pairwise comparison)
n = 17
hit rate = .88 (.33)
RT = 1684 (789)

SSIM
reference picture = 1.00
NoH-depiction = 0.86

NoH-distinction
n = 12
hit rate = 1.00 (0.00)
RT = 841 (686)

human-bird-distinction
n = 14
hit rate = .93 (.27)
RT = 722 (522)

NoH-distinction (pairwise comparison)
n = 16
hit rate = .94 (.25)
RT = 1552 (675)

SSIM
reference picture = 0.89
NoH-depiction = 0.88
“gap”

human; school age; boy

NoH-distinction
n = 59
hit rate = .59 (.50)
RT = 1201 (732)

human-bird-distinction
n = 61
hit rate = .93 (.25)
RT = 798 (524)

NoH-distinction (pairwise comparison)
n = 67
hit rate = .78 (.42)
RT = 1957 (781)

SSIM
reference picture = 1.00
NoH-depiction = 0.92

human; school age; girl

NoH-distinction
n = 64
hit rate = .66 (.48)
RT = 1360 (608)

human-bird-distinction
n = 65
hit rate = .88 (.33)
RT = 831 (495)

NoH-distinction (pairwise comparison)
n = 70
hit rate = .74 (.44)
RT = 2107 (847)

SSIM
reference picture = 0.99
NoH-depiction = 0.93

bird

NoH-distinction
n = 30
hit rate = 1.00 (0.00)
RT = 1173 (554)

human-bird-distinction
n = 71
hit rate = .92 (.28)
RT = 813 (458)

NoH-distinction (pairwise comparison)
n = 63
hit rate = .84 (.37)
RT = 2091 (756)

SSIM
reference picture = 0.94
NoH-depiction = 0.94
“shelf”

human; toddler; Caucasian

**NoH-distinction**
- \( n = 67 \)
- hit rate = 0.85 (.36)
- RT = 1138 (741)

**human-bird-distinction**
- \( n = 64 \)
- hit rate = 0.98 (.13)
- RT = 725 (436)

**NoH-distinction** (pairwise comparison)
- \( n = \)
- hit rate = 0.93 (.25)
- RT = 1616 (701)

**SSIM**
- reference picture = 1.00
- NoH-depiction = 0.82

human; toddler; African Native

**NoH-distinction**
- \( n = \)
- hit rate = (. .)
- RT = ()

**human-bird-distinction**
- \( n = \)
- hit rate = (. .)
- RT = ()

**NoH-distinction** (pairwise comparison)
- \( n = \)
- hit rate = 0.93 (.25)
- RT = 1658 (739)

**SSIM**
- reference picture = 0.65
- NoH-depiction = 0.73

bird

**NoH-distinction**
- \( n = \)
- hit rate = (. .)
- RT = ()

**human-bird-distinction**
- \( n = \)
- hit rate = (. .)
- RT = ()

**NoH-distinction** (pairwise comparison)
- \( n = \)
- hit rate = 0.88 (.33)
- RT = 1895 (845)

**SSIM**
- reference picture = 0.81
- NoH-depiction = 0.68
“shirt”

human; school age; boy

**NoH-distinction**  
\( n = 64 \)  
hit rate = 0.69 (.47)  
RT = 1438 (854)

**human-bird-distinction**  
\( n = 66 \)  
hit rate = 0.97 (.17)  
RT = 708 (413)

**NoH-distinction** (pairwise comparison)  
\( n = 70 \)  
hit rate = 0.89 (.32)  
RT = 1925 (796)

**SSIM**  
reference picture = 1.00  
NoH-depiction = 0.91

**bird**

**NoH-distinction**  
\( n = 41 \)  
hit rate = 1.00 (0.00)  
RT = 1311 (876)

**human-bird-distinction**  
\( n = 66 \)  
hit rate = 0.91 (.29)  
RT = 760 (437)

**NoH-distinction** (pairwise comparison)  
\( n = 67 \)  
hit rate = 0.88 (.33)  
RT = 2138 (837)

**SSIM**  
reference picture = 0.90  
NoH-depiction = 0.91
human; school age; girl

NoH-distinction
no data available yet

human-bird-distinction
no data available yet

NoH-distinction (pairwise comparison)
no data available yet

SSIM
reference picture =
NoH-depiction =
“sit”

human; toddler; boy

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 64</td>
<td>n = 68</td>
</tr>
<tr>
<td>hit rate = .72 (.45)</td>
<td>hit rate = .90 (.31)</td>
</tr>
<tr>
<td>RT = 1245 (723)</td>
<td>RT = 810 (520)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 66</td>
<td>n = 70</td>
</tr>
<tr>
<td>hit rate = .67 (.48)</td>
<td>hit rate = .93 (.26)</td>
</tr>
<tr>
<td>RT = 1478 (930)</td>
<td>RT = 695 (441)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction (pairwise comparison)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 71</td>
</tr>
<tr>
<td>hit rate = .97 (.17)</td>
</tr>
<tr>
<td>RT = 1945 (764)</td>
</tr>
</tbody>
</table>

SSIM

reference picture = 1.00
NoH-depiction = 0.78

bird

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 51</td>
<td>n = 64</td>
</tr>
<tr>
<td>hit rate = 1.00 (0.00)</td>
<td>hit rate = .95 (.21)</td>
</tr>
<tr>
<td>RT = 1228 (766)</td>
<td>RT = 811 (531)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction</th>
<th>human-bird-distinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 57</td>
<td>n = 63</td>
</tr>
<tr>
<td>hit rate = .70 (.46)</td>
<td>hit rate = .94 (.25)</td>
</tr>
<tr>
<td>RT = 1423 (866)</td>
<td>RT = 769 (525)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NoH-distinction (pairwise comparison)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 73</td>
</tr>
<tr>
<td>hit rate = .93 (.25)</td>
</tr>
<tr>
<td>RT = 1702 (703)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference picture = 0.84</td>
</tr>
<tr>
<td>NoH-depiction = 0.87</td>
</tr>
</tbody>
</table>
“stair”

human; kindergarden age; boy

NoH-distinction
n = 70
hit rate = .79 (.41)
RT = 1110 (634)

human-bird-distinction
n = 68
hit rate = .94 (.24)
RT = 646 (335)

NoH-distinction (pairwise comparison)
n = 71
hit rate = .87 (.34)
RT = 1815 (589)

SSIM
reference picture = 0.90
NoH-depiction = 0.79

human; kindergarden age; girl

NoH-distinction
n = 66
hit rate = .83 (.38)
RT = 1347 (841)

human-bird-distinction
n = 64
hit rate = .91 (.29)
RT = 668 (441)

NoH-distinction (pairwise comparison)
n = 70
hit rate = .87 (.34)
RT = 1889 (672)

SSIM
reference picture = 1.00
NoH-depiction = 0.96

bird

NoH-distinction
n = 49
hit rate = 1.00 (0.00)
RT = 1208 (786)

human-bird-distinction
n = 63
hit rate = .92 (.27)
RT = 738 (442)

NoH-distinction (pairwise comparison)
n = 69
hit rate = .91 (.28)
RT = 1961 (739)

SSIM
reference picture = 0.90
NoH-depiction = 0.96
“table”

human; toddler

| NoH-distinction | n = 53 | hit rate = .86 (.35) | RT = 1176 (934) |
| human-bird-distinction | n = 65 | hit rate = .91 (.29) | RT = 775 (417) |

| NoH-distinction | n = 63 | hit rate = .78 (.42) | RT = 1240 (725) |
| human-bird-distinction | n = 64 | hit rate = .94 (.24) | RT = 754 (399) |

NoH-distinction (pairwise comparison)

n = 71
hit rate = .96 (.20)
RT = 1519 (653)

SSIM

reference picture = 1.00
NoH-depiction = 0.88

bird

| NoH-distinction | n = 53 | hit rate = 1.00 (0.00) | RT = 961 (622) |
| human-bird-distinction | n = 70 | hit rate = .87 (.34) | RT = 739 (460) |

| NoH-distinction | n = 62 | hit rate = .74 (.44) | RT = 1185 (693) |
| human-bird-distinction | n = 63 | hit rate = .87 (.34) | RT = 745 (443) |

NoH-distinction (pairwise comparison)

n = 75
hit rate = .89 (.31)
RT = 1725 (706)

SSIM

reference picture = 0.86
NoH-depiction = 0.85
“table_chair”

human; kindergarten age; boy

NoH-distinction
n = 63
hit rate = .87 (.34)
RT = 1052 (676)

human-bird-distinction
n = 62
hit rate = .95 (.22)
RT = 719 (405)

NoH-distinction (pairwise comparison)
n = 74
hit rate = .96 (.20)
RT = 1588 (653)

SSIM
reference picture = 1.00
NoH-depiction = 0.81

bird

NoH-distinction
n = 63
hit rate = .86 (.35)
RT = 1250 (739)

human-bird-distinction
n = 66
hit rate = .89 (.31)
RT = 708 (457)

NoH-distinction (pairwise comparison)
n = 72
hit rate = .90 (.30)
RT = 1619 (655)

SSIM
reference picture = 0.84
NoH-depiction = 0.85

human; toddler

NoH-distinction
n = 67
hit rate = .88 (.33)
RT = 1043 (739)

human-bird-distinction
n = 66
hit rate = .97 (.17)
RT = 662 (349)

NoH-distinction (pairwise comparison)
n = 73
hit rate = .92 (.28)
RT = 1616 (683)

SSIM
reference picture = 0.78
NoH-depiction = 0.86
NoH-distinction
n = 73
hit rate = .92 (.28)
RT = 1616 (683)

SSIM
reference picture = 0.89
NoH-depiction = 0.84

NoH-distinction (pairwise comparison)
n = 74
hit rate = .96 (.20)
RT = 1563 (519)

human; kindergarten age; girl

NoH-distinction
no data available yet

human-bird-distinction
no data available yet

NoH-distinction (pairwise comparison)
no data available yet

SSIM
reference picture = 0.83
NoH-depiction = 0.87