Studying Cognitive Development in Infants and Toddlers

Jessica S. Horst  
University of Sussex  
jessica@sussex.ac.uk

Zoe M. Flack  
University of Sussex  
z.flack@sussex.ac.uk

Word Count  
5,114

Abstract  
Understanding the development of cognition in infants and toddlers informs many aspects of our understanding of human cognition. Designing research to address such questions with infants and toddlers provides a number of logistical and theoretical challenges. These challenges inspire innovative experimental methods and techniques. This entry outlines some of the key methods and innovations used to investigate early cognitive development in this special population. The entry also covers some of the research considerations and best practices to successfully study this fascinating and important topic.

Keywords  
Cognition, Development, Infants, Toddlers, Language, Categorization

Main Text

Conducting research on early childhood cognition is difficult. There are challenges that researchers in this area face that other cognitive scientists do not, such as short attention spans, participants crying, and parental interference. Consequently, researchers must balance their needs to create tests that robustly answer their research questions with the very real limitations of their young participants. For example, when investigating cross-situational word learning it is feasible to present adult participants with as many as 54 trials (e.g., Yu & Smith, 2007), but infants can only handle around 12 (e.g., Smith & Yu, 2008). Children’s abilities and limitations also change rapidly over the course of a few months, so some methods that are appropriate for young infants are now too easy only several months later. For example, detecting speech perception with high amplitude sucking is highly appropriate for newborn infants, but less so for 12-month-old infants. The following sections discuss some of the most common and insightful methods for testing cognitive development in infants and toddlers, respectively. This entry concludes with a summary of best practices.

Common Methods for Testing Infant Cognition

Habituation is a classic testing procedure for infants. Despite a long history in animal research, it was not until around 1970 that it was first used to study infant development (Cohen, 1976).
Habituation techniques have been commonly used in categorization studies (see Categorization in infancy, wecad00138). In this procedure infants are presented with a sequence of images (or videos) and the amount of time they look at these images is recorded. Stimuli are typically one at a time on a central screen in front of the infant. In some cases infants are shown one item repeatedly, in others they are shown multiple items across trials from the same category. With successive presentations infants’ looking time reduces. Once their looking time is past a threshold (typically less than 50% of the looking time they accumulated on the initial trials), they are considered to have habituated. At this point, a test phase begins where they are shown something new. If infants can detect a difference between this new stimulus and what they were exposed to before they should “recover” and look longer again, that is, they should dishabituate.

Habituation tasks are controlled by a computer, so that the operator is “blind” or unaware of the exact stimulus being displayed. This is achieved by having the operator only use the keyboard to indicate looking times and the computer automatically advancing the trials and also moving on to the test phase when the habituation criterion is reached. Computer programs like Habit 2 will also continue to a new trial when the infant looks away for a set amount of time that the researcher enters into the program when setting up the experiment. To regain infants’ attention to the screen at the start of the next trial an “attention getter” is displayed. The best attention getters are dynamic and include sound. Common attention getters include faces of giggling babies and shapes that loom with fun noises like chimes or bicycle bells. The computer operator is often sitting behind a curtain and observing the infant through closed-circuit television (CCTV), which also ensures he/she does not see the displayed stimulus.

Habituation procedures can use a fixed window, where researchers take the amount of time infants look at the stimulus in a set of trials, for example, trials 5-8, and compare that to their looking in the initial window (i.e., trials 1-4). If the looking time on trials 5-8 is larger than 50% of the looking in the initial window, then trials 9-12 are presented and compared to the initial window, and so forth. Another option is a sliding window, for instance comparing trials 5-8 to the initial window, then trials 6-9, then trials 7-10 and so forth.

It is highly common in habituation procedures for different participants to have different amounts of cumulative looking times. An infant that habituates after 8 trials will typically have viewed the stimuli for less time than an infant who does not habituate until 16 trials. Importantly, though, it is the infant’s own processing speed that determines when the test phase will begin. In some cases researchers prefer familiarization procedures in which the test trial commences after a set number of trials or a pre-determined “cumulative looking criterion” is reached. These procedures ensure infants reach the test phase with very similar amount of experience, but do not control for individual differences in encoding and processing.

Preferential looking, and intermodal preferential looking (IPL), are similar to habituation in that infants are typically seated in some kind of booth and look straight ahead. “Preferential” refers to the infant’s choice of what to look at: the left or right stimulus. “Intermodal preferential looking” indicates that infants are both seeing and hearing stimuli, for example, seeing pictures and hearing their names. Preferential looking studies often present infants with items that differ in visual features such as color or different items from the same category or two objects and one label. The two stimuli can be shown on a large screen or on two screens with a light between the screens serving as an attention getter.
Unlike habituation procedures, preferential looking procedures present infants with a set number of trials. Further, rather than comparing infant looking times across trials, in this task infants’ preferences are usually calculated as the amount of time they spent looking at one item out of the total amount of time they spent looking at all of the items. Importantly, this kind of calculation takes into account that sometimes infants will look elsewhere such as at their own shoes rather than at the stimulus. So for example, using this calculation the researcher can determine infants’ “novelty preference,” which would be the amount of time they looked at a novel item out of the amount of time they looked at both a novel and a familiar item.

Some studies that use the intermodal preferential looking procedure employ the “switch task” (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). In this task, infants or toddlers are presented with pairs of visual and auditory stimuli. For example, infants might see Object A with Sound A and Object B with Sound B. At test, the items are switched, so in this example infants would see Object A with Sound B. If infants have learned the intermodal association they should look longer to this incorrect pair than to a correct pair (Object A, Sound A). Switch tasks are common in studies on early language processing, but are not exclusive to language studies.

The violation of expectancy paradigm analyzes infants’ increased looking times in response to unexpected events (see Visual perception in infancy, wecad00110 for further discussion). During training children typically watch two alternating events and then at test are shown something that would appear impossible to adults. Such events typically involve an object moving behind an occluder (e.g., screen) or behind another object. For example, previous studies have shown infants events where a rabbit moved along a track and behind an occluder that contained a window in the middle. In one event the rabbit was small and should not have been tall enough to be seen through a window (expected), and in the other the rabbit was taller and should have been seen through the window as it passed, but was not (unexpected). Children looked longer at the unexpected event than the expected event, which was interpreted as evidence they understood this is an impossibility and it violated their expectation. Violation of expectancy is often used to examine infants’ understanding of the physical world and simple arithmetic, although others have argued that lower-level perceptual and attentional aspects of the task could be driving the effects (e.g., Bogartz, Shinskey, & Schilling, 2000).

Infants’ preferences for certain stimuli can be used to study cognitive development using other methods as well. For example, both high amplitude sucking and head-turn preference paradigms make use of children’s preference to hear a particular stimulus, such as their native language or familiar words. As in preferential looking studies, this preference can be used to test if children can make distinctions between stimuli. In high amplitude sucking an initial training procedure uses operant conditioning to associate auditory stimuli with sucking on a specially-designed pacifier/soother. Infants’ sucking amplitude then reflects their preference for the auditory stimulus: they suck more to keep hearing the auditory stimuli they prefer.

The head-turn preference procedure can be used in a similar way. Researchers can measure the direction of the head turn or the duration of the head turn. Typically, infants are trained to look in a specific direction to a toy that is activated when a particular stimulus is played. Over time, the delay before the toy is activated is increased so that the researchers can measure anticipatory looking. In these tasks an experimenter serves as a kind of “attention keeper”: maintaining infant’s looking straight ahead by showing the infant toys. The
experimenter should wear headphones to remain “blind” to when the infant should turn his/her head.

Looking-while-listening is a method related to preferential looking, in which young children, often toddlers, are presented with pairs of pictures, but with accompanying speech naming one of the pictures. Meticulous frame-by-frame coding of the infant’s looking patterns identifies when exactly children look at the pictures. This rich data requires considerable effort to clean, code and analyze, making it rather time-consuming. However, the data show the time course of emerging comprehension in minute detail, which enables fine-grained analysis with remarkable precision. For a full description of this method, see Fernald, Zangl, Portillo, and Marchman (2008).

Eye tracking is now widely used because it also enables fine-grained analyses of how learning unfolds over time, but is much less labor-intensive to code. Early eye trackers typically required a participants’ heads to remain in a fixed position, which made them unsuitable for children. However, growing commercial interest in this area, for gaming and mobile technology, as well as interest for research purposes, has led to the development of increasingly flexible solutions. For example, a variety of desktop eye-trackers can be used to track eye movements on a screen and head-mounted or wearable eye trackers such as glasses can be used to track eye movements in the real world. Most studies using eye-tracking techniques seat infants and toddlers in front of a computer screen, where images and videos are displayed. However, head-mounted eye-trackers are also used to get a better sense of what children view during their everyday activities (e.g., Franchak, Kretch, Soska, & Adolph, 2011).

Eye trackers thus provide extraordinarily rich data, which can be used to identify more than just where children look. Fixations, saccades and pupil size can all be used to investigate aspects of children’s developing cognition such as vision, attention and perception. Stimuli can be presented on a screen either as static or dynamic imagery. Particular areas of interest, or periods of interest can be specified during which looking measures can be collected. For example, if children were presented with two stimuli the researcher could identify each as an area of interest, and compare looking times. When accompanying audio is provided, the researcher may only be interested in looking behavior after the onset of the audio, so could specify a time window. Eye tracking studies can be affected by certain issues. For example, inaccuracies in the data can be introduced if equipment is poorly calibrated or unsuited to the planned research. Excessive movement by children can easily cause lapses in eye movement recording, which can then result in quite substantial windows of missing data. In addition the vast data captured by eye-trackers present a number of potential analytical challenges for the researcher, making the definition of a clear research objective at the outset particularly important.

Other methods have been used to investigate the brain processes associated with cognitive processing. Electrophysiological methods such as electroencephalography (EEG) can be used to measure changes in the electrical activity of the brain by placing tiny electrodes on the scalp. Using this method an ‘event related potential’ (ERP; a response to stimuli), can be measured, allowing us to understand a great deal about the timing of cognitive processes. Another method for examining cognitive development is near-infrared spectroscopy (NIRS). This method has advantages for infant and toddler research over more traditional imaging methods such as fMRI (functional Magnetic Resonance Imaging) and EEG/ERP. For example,
NIRS is cheaper, less invasive, portable, and considered to be safer. In addition, NIRS is less perturbed by movement than other methods which makes it particularly suitable for use with infants and toddlers. Near-infrared light is used to measure changes in oxygenation in hemoglobin levels, which in turn can be used to map the brain with a good level of both spatial and temporal resolution (Minagawa-Kawai, Mori, Hebden, & Dupoux, 2008). Since its arrival, NIRS has been used to investigate infant cognitive development in topics such as face, object and language processing.

**Common Methods for Testing Toddler Cognition**

As children age, experimental tasks begin to introduce manual responses by asking children to point to, take or simply play with objects. The object examining task is a very similar to preferential looking tasks in design, but allows children to be more active (Oakes, Madole, & Cohen, 1991). In this task, infants are presented with toys to look at and examine, for example dogs and horses (Kovack-Lesh & Oakes, 2007), rather than images. Researchers review video footage to measure the amount of time infants spend looking at and engaging with the objects. They take one pass of the video footage to measure the examination time for the first object. Then they take another pass of the footage for the second object. In this way, novelty preference or other scores can be calculated as they are in preferential looking and familiarization tasks.

Sequential touching is a similar method except that toddlers are presented with several items at once, typically eight; four objects from each of two categories, for a short period of time, e.g., 2 minutes. For example, infants may be presented with animals and vehicles, balls and blocks or dogs and horses. For this task, the order or sequence in which objects are touched is the critical measure, and from this sequence mean run lengths (MRL) are calculated. A MRL is calculated for each categorical contrast (e.g., animals versus vehicles) by dividing the total number of touches by the number of “runs” or sequences in which all of the touches are to the same category (e.g., three touches to animals would be a run of 3). The MRL is then compared to a level of chance, unique to each total number of touches, based on Monte Carlo simulations (Mandler, Fivush, & Reznick, 1987).

The difficulty with sequential touching studies is that they are highly laborious to code; it takes on average 45 minutes to code 10 minutes of sequential touching footage. In addition, like habituation, children in the same study may end with vastly different levels of experience, some toddlers may touch only some of the objects while others may touch all of objects, some toddlers may make very few touches (e.g., 5) while others make many (e.g., 30). Objects are routinely dropped and thrown in this task and the researchers must reach a consensus on whether or not to include such touches and how to score them.

Looking tasks including habituation and preferential looking, object examining or sequential touching tasks can lead to high attrition rates when testing toddlers. This high attrition rate suggests these tasks may not be the most appropriate testing methods for toddlers. Many studies with toddlers include forced-choice trials with objects or pictures. On forced-choice trials children are presented with two or more objects or pictures and they are asked to choose one of them by choosing the object they like the best, choosing the object that is most like another item or choosing the object that they are in the process of associating with a particular name (e.g., “get the blicket”). Forced-choice trials are especially common in studies that examine mutual exclusivity and word learning biases. They are highly enjoyable and children complete them
effortlessly, however, because of the ease of these tasks for children, we are at risk of overestimating children’s competency.

Critically, the exact wording that experimenters use in these kind of tasks can matter, for example, it is easier for children to grab an object and keep it for themselves, than to pass the object back to the experimenter—especially when the object is something novel and exciting to touch and explore themselves. So in this case, asking them to, for example, “get the____”, is easier than for example, “give me the____”. In addition, the kinds of syntax used also influence infants’ behavior in the task, as can the number of objects to choose from and even the infant’s posture.

Once children are a little older, and can nod or shake their head, or give yes or no responses, it is possible to ask children yes/no questions, for example to show them an object and ask “is this a ____?” to allow children to indicate whether they believe the object is called by that name. However, for young children the data obtained from yes/no tasks differs from the data from forced-choice tasks (Samuelson, Schutte, & Horst, 2009) and children under 4 years of age are biased to answer yes (Heather Fritzley & Lee, 2003). Language comprehension precedes production (Caselli, Rinaldi, Stefanini, & Volterra, 2012) and it can be difficult for children to articulate their responses. Studies that ask children to produce novel words often end up with close to floor level responses.

Because production and verbal responses are more difficult than simply choosing objects, researchers have recently developed the dot task (originally called the spatially-supported name recognition task (Gordon & McGregor, 2014)). In this task children are presented with a piece of paper that includes a series of dots and the experimenter points to each dot in turn and says a corresponding name such as dorb or zinnip. Then children are asked to point to the name of an object and point to the dot that was associated with that name. This task allows children to respond based on recognition rather than recall, which is more comparable to other tasks.

Research Considerations

Like all psychological experiments, the precise stimuli used matter in research on early cognitive development. Researchers have found relatively low-level differences in stimuli can have profound effects on children’s responses, for example, changes in vowel sounds, background color or inter-trial interval. In addition, infants’ and toddlers’ test performance is highly susceptible to interference from previous experience. For example, children’s performance may change depending on what they chose on previous trials, the order in which the items were presented, pre-exposure to the items, and any exposure they may have had at home. Thus, care should be taken to control for stimuli, exposure and presentation. Overall, stimuli should also be chosen specifically for the target age group and real objects should be robust enough to withstand rough handling by young children. Because many children might handle the objects, and because children do occasionally put objects in their mouths, it is also useful to consider how objects will be cleaned between experimental sessions.

Infant preference underpins several methods for investigating cognitive development. However, infants cannot demonstrate a preference if they cannot discriminate between two items (e.g., if they cannot see that two different animals are not the same). For this reason, it is often advantageous to first run a discrimination study before the main experience to ensure infants can
discriminate between the stimuli. Infants have been shown to demonstrate either a preference for novelty (e.g., dishabituation) or a preference for familiarity. In many research designs either preference is acceptable; having a preference indicates an understanding that the two stimuli are different. Researchers argue that infants demonstrate a familiarity preference when they have not fully processed the stimuli. That is, as they continue to process something they continue to look at it. Once they have fully processed the stimulus, then they switch and demonstrate a novelty preference. This switch explains the individual differences and variability often found in infant looking-time data. Infants can fail to dishabituate because either they did not notice a change or they are fatigued, therefore many researchers include a “completely novel” test trial as the final trial in their looking studies (e.g., a vehicle after a study about cats and dogs). If infants fail to dishabituate and do not “recover” for this final trial, this pattern of behavior suggests they did not dishabituate because of fatigue. In contrast, if infants fail to dishabituate and do recover for this final trial, this pattern of behavior suggests they did not notice the difference between the test stimulus and the stimuli presented earlier. Toddlers often demonstrate a novelty bias (e.g., Mather & Plunkett, 2012; Horst et al., 2011), so stimuli need to be well-controlled. In addition, control trials with familiar items are often included to ensure effects are due to the experimental manipulation and not toddler’s intrinsic novelty bias.

Best Practices

Studying cognitive development in infants and toddlers is expensive, both in terms of time and money. Arranging to visit children in their homes or preschools as well as asking parents to bring children to labs involves considerable preparation and payment for transport costs. In addition, although individual experiments may only last a few minutes due to short attention spans, additional time needs to be set aside before testing to ensure children are comfortable interacting with the researcher. As a result, sample sizes can be smaller than those found in research conducted with adults. This can result in low powered experiments, which in turn can lead to both false positive and false negative results. Oakes (2017) provides a discussion of the issue and offers some solutions for the field.

Another reason research with children is expensive is because higher levels of children than adults typically fail to complete experiments. This attrition is often listed in academic publications as “fussiness” and may be due to crying, falling sleep, shyness or not engaging with the experiment in some other way. High attrition rates are not related to obtaining positive results (Slaughter & Suddendorf, 2007), but is one reason more extensive recruitment is required for developmental research than adult research. Depending on age and the method used, attrition rates can vary greatly. For example in ERP studies where excessive movement can be problematic, attrition rates can be as high as 25%-75% and this research becomes particularly difficult with children after the first year (Hoehl & Wahl, 2012). In infant behavioral studies, it is more typical for 20% or fewer of the participants to be excluded.

Recruiting participants can be particularly challenging for those studying cognitive development. Although the advent of social media has made targeted advertising relatively easy and affordable, there are a number of obstacles, which prevent parents from coming in to a lab.Ads need to engage parents’ attention quickly, and at the right time, so consideration should be given to how and when ads are delivered. In addition, parents of young children often feel strapped for time, so any sign-up forms need to be brief and only ask a few questions. Local
daycares, playgroups, community centers or similar are often happy to host visitors, which can enable access to suitable populations and may even allow for experiments to take place on the premises, which can be very efficient for some types of data collection. However, if parents need to visit a lab, then their travel costs may need to be reimbursed. It is also customary to give children a small gift for participating, such as a cuddly toy or storybook, though some labs also use logoed t-shirts, which can also spread the word about the lab.

When scheduling experimental sessions, it is important to proactively ask parents to consider their child’s routine as children who have missed a nap or just been swimming may be irritable and uncooperative, meaning a wasted journey and a negative experience for all. If working in daycare settings or preschools care should be taken to avoid causing too much disruption to the established routines, and some consideration given to the impact of the inevitable distractions in such busy environments.

For lab visits, infants and toddlers are always accompanied by a caregiver, which introduces additional methodological considerations. Parents are invested in the success of their children and are often keen to see their children perform well in experiments. Parents can even unknowingly provide cues, which could influence the validity of data. It is important to ensure experimenters strike a balance between involving parents and protecting experimental control and validity. This can be achieved by observing a few key guidelines. Parents should be provided with enough information to give informed consent, so the experimenter should brief parents on what will be required of the child, without providing explicit hypotheses. It can be helpful to politely explain to parents the importance of allowing children to complete tasks without their support to avoid any unplanned interference. For language studies, it is often helpful to tell the parent what they are or are not allowed to say (e.g., “if you need to name a toy, call it what I call it” or “you can talk about the toys, but don’t name them”). It may even be necessary to provide parents with black-out glasses or headphones to ensure they cannot see or hear stimuli. Researchers who use such glasses and headphones often playback the stimuli for the parent after the experiment is completed (doing so before could provide pre-exposure to the child).

Once parents have been prepared, children should be provided with age-appropriate instructions. Instructions should be unambiguous, positive and avoid unnecessary questions. Care should be taken to ensure children are encouraged to participate, without compromising their ability to withdraw from the experiment if they wish. For example, asking a toddler “would you like to play with some toys?” is encouraging and informative for that age-group. Infants and toddlers may struggle to articulate their desire to withdraw as participants, so experimenters should err on the side of allowing their young participants to withdraw if they appear distressed. Again, attrition rates are often very high in studies with young children, but they are not indicative of the overall quality of the data (Slaughter & Suddendorf, 2007).

Inter-coder reliability is highly important because some methods for testing infants and toddlers are arguably subjective because they are coded by humans. The convention is to select a random sample of 20% of the data to be coded by a second coder, who is ideally blind to the experimental hypotheses or even to which stimuli are the target stimuli. Experiments with young children are often video-recorded not only for the initial coding, but also for coding such reliabilities. In many cases a second coder codes the footage independently and then the scores are compared. For studies in which responses are categorical and the same coders code all sessions, Cohen’s Kappa can be calculated as a measure of inter-coder reliability that also takes
into account agreement occurring by chance. However, in some cases the assumptions of Cohen’s Kappa cannot be met so other measures of inter-coder reliability are used. For example, in some looking studies the looking times per trial are often correlated but also the mean difference is also reported as coders could have high correlations but consistently disagree on how long the infant looked. For forced-choice trials and sequential touching tasks inter-coder reliability is calculated as the number of times the coders agreed divided by the number of times they could have agreed (which is usually the number of trials a child answered). It is good practice to code for inter-coder reliabilities regularly over the course of the experiment and not at the very end. For example, to have a routine of coding one reliability after every 5 children are tested (but still randomly selecting which sessions to code). Coding over the course of the experiment can allow the team to catch any experimenter bias or errors before the full sample is tested. It can also help catch situations of “coder drift,” when a coder becomes less reliable over time due to idiosyncrasies.

The issues discussed here outline some of the challenges of working with infants and toddlers. Despite these challenges, this kind of research is highly rewarding. It is fascinating to what a young mind learn something new and to even see that learning unfold in front of your very eyes.

SEE ALSO:
Basic research (wecad00055), Categorization in infancy (wecad00138), History of the behavior analysis of child development (wecad00049), Learning in infancy (wecad00140).

References
Further Readings


Author Biographies

Jessica S. Horst (Ph.D., The University of Iowa, 2007) is a developmental psychologist and reader at The University of Sussex and Director of the WORD Lab. Her research focuses on cognitive and perceptual factors that support young children’s memories for newly learned words. Her basic research methods have inspired community interventions that successfully promote word learning from repeated storybook reading. She is an editor of The International Handbook of Language Acquisition (2019, Routledge) and author of The Psychology Research Companion: From Student Project to Working Life (2015, Routledge), along with dozens of empirical articles on early language acquisition and categorization.
Zoe M. Flack (Ph.D., The University of Sussex, 2018) is a developmental psychologist and lecturer at The University of Brighton. Her research interests include education, particularly education which takes place within the home environment. Prior research has examined learning from shared reading encounters, and current projects consider ethical and methodological issues associated with researching learning in the home. She is the author of The Effects of Shared Storybook Reading on Word Learning: A Meta-Analysis, Developmental Psychology (2018).