
Volume 29, Number 2, March/April 2008
ISSN 0891-4222
ELSEVIER
Research in Developmental Disabilities

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Research in Developmental Disabilities 29 (2008) 149–157
Research in Developmental Disabilities

Mapping images to objects by young adults with cognitive disabilities

Stefan Carmien [a,*], Erica Wohldman [b]

[a] *Fraunhofer Institute for Applied Information Technology, Schloss Birlinghoven, D-53757 Sankt Augustin, Germany*
[b] *Department of Psychology, California State University, Northridge, CA, USA*

Received 31 January 2007; received in revised form 8 February 2007; accepted 13 February 2007

Abstract

How the type of representation (icons, photos of objects in context, photos of objects in isolation) displayed on a hand-held computer affected recognition performance in young adults with cognitive disabilities was examined. Participants were required to match an object displayed on the computer to one of three pictures projected onto a screen. We tested the opinion widely held by occupational therapists and special education professionals that there is an inverse relationship between cognitive ability and the required fidelity of a representation for a successful match between a representation and an external object. Despite their widespread use in most learning tools developed for persons with cognitive disabilities, our results suggest that icons are poor substitutes for realistic representations.

Keywords: Image recognition; Developmental disability; Iconicity

The Cognitive Leavers research project at the University of Colorado Center for Lifelong Learning and Design (L3D) develops socio-technical environments to support independence for persons with cognitive disabilities and their caregivers. One of their projects, the Memory Aiding Prompting System (MAPS), is a hand-held prompting system (Carmien, 2004a) that prompts a person with cognitive disabilities through tasks that were previously difficult or impossible for them to complete independently due to difficulties with memory or executive functions. The prompts consist of an image and a verbal instruction; a script consists of a sequential series of these prompts. Scripts are typically made by caregivers, and are specific to the user (i.e., the person with cognitive disabilities) and the task (Carmien, 2005). In the process of developing MAPS, it became clear that the type of image displayed on the screen of the hand-held computer

* Tel.: +49 2241 142639.
E-mail address: stefan.carmien@fit.fraunhofer.de (S. Carmien).

0891-4222/$ – see front matter
doi:10.1016/j.ridd.2007.02.003

could affect strongly the success of the attempted task. The images used by computationally based augmentative and alternative communication systems (AAC) (Beukelman & Mirenda, 1998) and prompting task support systems can be a weak link in providing assistive technological support for persons with disabilities that require their use. When an image is not instantly recognizable as referencing an object in the world, then there are immediate consequences to the person attempting to use it successfully to accomplish a given task.

The two classes of systems that use images as supports, AAC and prompting support, have similar goals but differ with respect to how the image is used. Specifically, AAC images are intended to replace or supplant communicative ideas, which can range from simply pointing to an object in the world (''a loaf of bread'') to encompassing an entire concept (''buy a movie ticket'') (Beukelman & Mirenda, 1998). Computationally supported multimedia task support systems (Davies, Stock, & Wehmeyer, 2002), on the other hand, have the ability to use a recording of a verbal prompt to provide the 'verb' required to perform the task. Because of this difference, the emphasis of the present study was to determine the type of image necessary for successful identification of an object in the world. The target population for this experiment were young adults with cognitive disabilities, which is defined by Diagnostic and Statistical Manual of Mental Disorders IV (American Psychiatric Association. Task Force on DSM-IV, 2000) as a person who is ''significantly limited in at least two of the following areas: self-care, communication, home living social/interpersonal skills, self-direction, use of community resources, functional academic skills, work, leisure, health and safety.'' Four different degrees of cognitive disability are defined: mild, moderate, severe, and profound. In our experiment, we included only persons with mild cognitive disabilities (IQ 50–55 to 70) and those in the upper range of moderately disabled (IQ 35–40 to 50–55). We focused on young adults because in this particular geographic region (Boulder, Colorado) research has shown that young adults (13–24) with mild to moderate cognitive disabilities possess the skills to use and understand simple computers such as cell phones, TV remote controls, etc., unlike their elders (Carmien, 2006). Thus, these participants had very few problems operating the simple handheld computers used in our tests.

There have been only a limited number of studies published on the topic of how best to represent real-world objects for individuals with cognitive disabilities, which is the primary goal of this experiment. In the AAC world, there have been numerous studies on the quality of iconicity, transparency and guessability of images. Wilkinson and McIlvane (2002) provide a good overview of these topics, defining iconicity as ''used to describe the visual similarities'' and as ''the relationship between a symbol and its referent.'' Iconicity is not an absolute value of an image, rather it is a characteristic of the observer (Stephenson & Linfoot, 1996). Biederman's (1995) research on visual object recognition discusses the fundamental problem of object recognition from hierarchical levels of classification. For example, a bird could be represented as an animal, a bird, a parrot, or a specific bird named Pete. The entry-level classification determines the success of matching an image with its object in the world. For example, tomato sauce does not necessarily have to be depicted as a specific brand of sauce, but must be represented as more than a can with a red label, an issue of the level of classification (Biederman, 1995) of the object by the perceiver.

Another approach to understanding how best to represent objects was taken by Dixon (1981), who studied how life-sized photos were matched to objects for children with severe intellectual disabilities. Dixon's results suggested that the physical properties of the objects in a photo (e.g., shape, surface texture, and gradient) and the depth dimensions of the object in the photo are more predictive of whether or not a successful match will take place than the similarities of the objects and images within the photos. Mirenda and Locke (1989) studied the mapping of symbols (e.g., icons, pictures, 3D models, and words) onto real objects in non-speakers with varying levels of cognitive disabilities, and found that 3D models were easier than any other symbols to mach.

Furthermore (Stephenson & Linfoot, 1996) reviewed the existing literature on picture recognition in persons with cognitive disabilities, and how pictures come to be used as symbols for the objects they depict. The authors, in their survey of research, discriminate between picture recognition and picture use, picture use being pictures utilized as communicative support. The studies reviewed supported the notion that picture matching with objects is mastered before the use of images to communicate, and the matching skill is separate from the communicative skill. Several studies support the position that images out of context, or, more properly in no context (on a white background) are more conductive to object matching. Other studies emphasized that picture recognition is a learned skill (as well as being developmentally based) and in support of this discussed various studies of different cultures approach to object representation and matching.

A common ''rule of thumb'' used by assistive technologists is that in order to obtain a match between a representation and the real-world object being depicted by the representation, the image fidelity must be inversely related to the level of cognitive ability. That is, lower abilities result in a greater need for fidelity of the image (e.g., Snell, 1987). We refer to the formulation of this conjecture as Carmien's rule. In Carmien's rule, the term 'image fidelity' refers to a hierarchy of representations, from its actual physical form, to a model or photograph of the actual object, to a photograph of a specific brand of object, and finally to written words. It is not identical to iconicity or guessability, as the iconicity of a representation is a relationship between the observer, the representation, and the 'target' object (e.g., Wilkinson & Mcllvane, 2002). While research has pointed out that the iconicity of a representation is an idiosyncratic phenomenon, at least in specific instances, the authors posit that the need for higher fidelity is only loosely connected with cognitive ability—varying not only from person to person but also possibly from culture to culture (Huer, 2000).

While there have been many studies that indicate that the inspection time[1] to recognize an image is inversely correlated with IQ (Mackintosh & Bennett, 2002), none of the previous studies demonstrate conclusively Carmien's rule. There are pragmatic tools that have been designed to very generally determine viewing preferences for specific individuals with cognitive disabilities (Assistive Technology, 2005). In addition, some books teaching best practices also discuss the issue of how a representation is recognized in persons with cognitive disabilities (Snell, 1987). However, we wanted to test directly the effects of the type of picture cue provided to persons with cognitive disabilities on the success of matching the image to the external object that it represents. Thus, the primary question being addressed by this experiment involved whether the type of images could influence matching performance.

1. Method

1.1. Participants

Fifteen high-school students (nine males, six females, $m = 15.8$ years old) from the special education classes in the Boulder Valley School District were invited to participate in this

[1] Visual inspection time (IT) is defined as the minimum time an individual requires to reach near perfect response accuracy in a simple two-choice visual discrimination.

experiment. Informed consent from both the student and his or her legal guardian was obtained. Students were selected on the basis of their IQ scores and their ability to use the MAPS script guidance system, as well as on their need for such a mnemonic and executive cognitive othontic to accomplish activities of daily living (ADLs) independently. There was a wide range of cognitive disabilities, including but not limited to autism, downs syndrome, alcohol related neurological disorders, and fetal alcohol syndrome. IQs (using the WAIS scale) ranged from 43 to 85 ($m = 58.1$). In addition, 15 age-matched high-school students (nine males, six females, $m = 15.9$ years old) with no cognitive disabilities were selected from classes in the Boulder Valley School District.

1.2. Materials and apparatus

A Compaq model 900c laptop computer was used to test all participants in this experiment. The program was written using Visual Basic V6 and Embedded Visual Basic. A Proxima 5900 projector displayed three target images onto a white screen. The computer was connected to a wireless router that established a local area network (LAN), which was connected to a hand-held Compaq IPAC model 3670 computer, with a two and one quarter by four inch screen. There were three large push buttons (6 in. in diameter), in three colors (yellow, red, and green), attached to the laptop.

1.3. Procedure

Participants were tested individually in a quiet room in either the L3D lab or the high school at which the participant attended. A trial consisted of matching a picture displayed on the hand-held computer to one of three pictures displayed on the white screen by pressing one of three buttons that were placed in front of each picture on the white screen. Response time was measured as the time that participants took to press one of the three buttons after the onset of the image on the white screen.

Each image (which we will call the cue) was scaled to fit on the hand-held computer (280×210 pixels), and the images displayed on the white screen included the correct target, a target lure (an item that could possibly make for a suitable substitute in place of the target), and non-target item (an item that was neither the target nor a near substitute for the target, but had a similar shape or size to the target). For example, if the target image was jar of spaghetti sauce, then the target lure might be a jar of salsa and the non-target might be a jar of peanut butter). Each target could be represented on the hand-held computer in one of three ways. Specifically, the picture was either shown in isolation (a single target object shot against a white background), in context (a view of the target item on a grocery shelf), or was represented as an icon from the set of Mayer-Johnson Picture-symbol icons (Mayer-Johnson, 2001) that are typically used in instruction for this specific population of young adults.

A short break was given after every 10 trials were completed, and for each participant the experiment consisted of two blocks of 30 trials with 10 of each picture type (icon, isolated, context) in a block. Each block had 10 set of targets paired with one of the three kinds of cue images on the PDA. The 30 trials were randomly scattered across a block. In this way every participant was presented with every kind of cue for a given target. The images were stored on a Sybase database on both the hand-held and laptop. Each image (2 ft $\times$ 3 ft) on the white screen was surrounded with a border that was the same color as the button located underneath it (see Fig. 1).

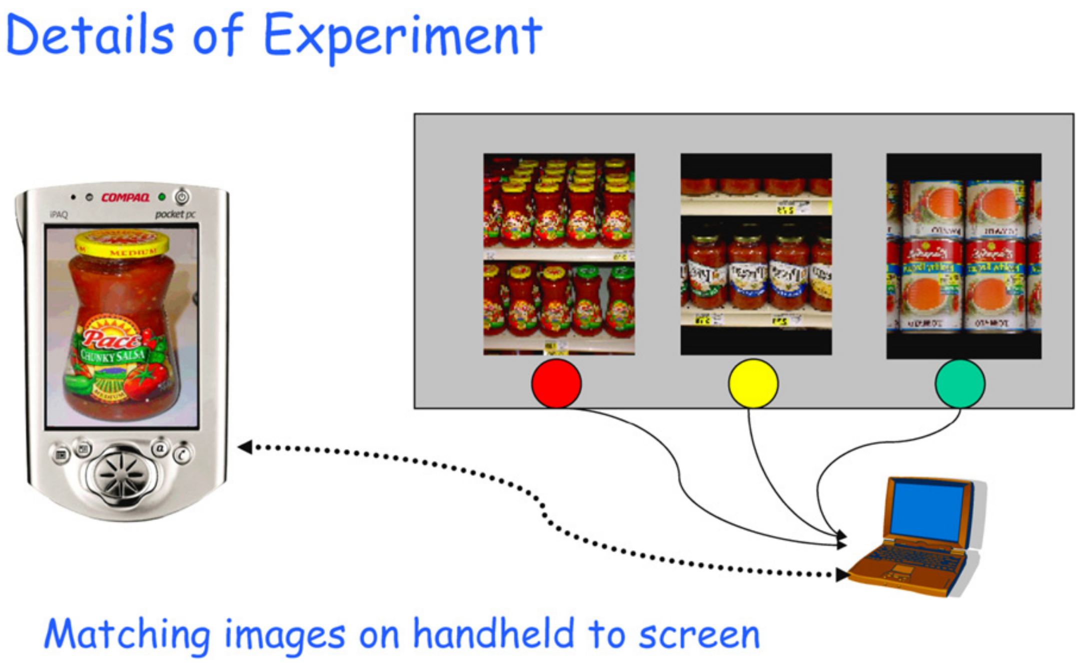

Fig. 1. Block diagram of experimental setup.

The experiment began with a demonstration to show how the task was to be performed. The demonstration consisted 10 trials, first showing the participant how to match the hand-held image to one of the three possible targets shown on the screen by pushing the appropriate button. Next, the participant went through the trials during which verbal cues were provided by the experimenter before finally practicing the remaining trials without assistance. After this initial practice phase, the experimenter imported randomly selected pictures onto the hand-held computer. The hand-held computer had two states—currently showing the target image and ready to show the next target. The participant activated these states by touching the screen on the hand-held computer. Once the set of pictures had been selected, the hand-held showed the 'ready' image (see Fig. 2). The participant was then asked to touch the screen when ready and the hand-held transmitted a signal to the laptop, triggering the laptop to both import the images that the participant would see on the projector screen and to send a signal back to the hand-held to display the target image on its screen. The images appeared simultaneously on the hand-held computer screen and on the white screen. When the participant pressed one of the three buttons, the computer obtained information identifying the image and a code that indicated whether that image was a perfect match, a lure match, or a nontarget match.

1.4. Design

Two dependent measures were included in our analyses. First, response time was calculated as the time that it took participants to press one of the three buttons after the pictures were displayed on the white screen. Only accurate trials were included in the response time analyses. Second, accuracy of the response was calculated and reported as proportion correct. The overall design for this experiment was a mixed factorial including participant type (cognitively disabled, and cognitively typical) as a between-subjects variable and block (first, and second) and target type on the hand-held computer (isolation, context, and icon) as within-subjects variables.

Fig. 2. Hand-held with ready image.

2. Results

A repeated measures ANOVA was conducted on both response time and accuracy. Two types of analyses were conducted on each measure. In the first type, we included all three-target types (isolation, context, and icon) for comparison. In the second type, we compared icons to the average of realistic images (i.e., to the average of those in isolation and in context) to determine how suitable icons are for representing real-world items. One typical participant was removed from the analysis because his accuracy performance was more than four standard deviations (.102) away from the mean of his group.

2.1. Response time

Both the first and second analyses of response time yielded a significant main effect of participant type, $F(1, 27) = 15.48$, MSE = .874, $p < .001$, $F(1, 27) = 15.53$, MSE = .578, $p < .001$, respectively. Typical participants (first = 2.30 s, second = 2.39 s) showed faster response times compared to cognitively disabled participants (first = 7.88 s, second = 7.96 s). No effects involving picture type were significant.

2.2. Proportion correct

In both types of analyses conducted, partially correct (i.e., the selection of a target lure) trials were included because target lures would make a suitable substitute for the target item and because there were very few correct trials from the participants with cognitive disabilities. Thus, correct trials were given a score of 1, correct target-lures were scored as .5, and non-targets were scored as 0.

As expected, the first type of analysis yielded a main effect of participant type, $F(1, 27) = 31.21$, MSE < .075, $p < .001$. Specifically, typical participants (.960) were more accurate on the matching task compared to those with cognitive disabilities (.727). In addition, participants were more accurate when the object was represented realistically either in context (.876) or in isolation (.862) than when it was represented as an icon (.779); the main effect of picture type was significant $F(2, 54) = 12.97$, MSE = .012, $p < .001$.

In the second analysis comparing icons to realistic images, the main effect of participant type was also significant, $F(1, 27) = 37.49$, MSE = .046, $p < .001$. Again, typical participants (.951) were more accurate than those with cognitive disabilities (.706). In addition, the main effect of picture type (icon, and realistic) was significant, $F(1, 27) = 19.52$, MSE = .012, $p < .001$, with, overall, higher accuracy on realistic images (.869) compared to icons (.779). Furthermore, the interaction of participant type and picture type was marginally significant, $F(1, 27) = 3.35$, MSE = .012, $p = .078$. Participants with cognitive disabilities were less accurate when the pictures were represented as icons (.643) than when they were represented realistically (.768), and although typical participants showed a similar trend, the difference in accuracy between icons (.925) and realistic images (.977) was smaller. This finding suggests that icons may be poor substitutes for realistic representations, especially for people with cognitive disabilities.

3. General discussion

Although the results show that for both typical and cognitively disabled young adults icons are less likely to lead to a quick and accurate target selection than are realistic images, there was no significant difference between the two types of realistic images (isolated and in context). This finding was unexpected, as existing research (Braun, 2003) tends broadly to emphasize context and contextual cues as significant in object recognition. There are at least two possible explanations for our results. First, the images we called ‘in context’ might not be in context in the way that the other studies used the word. Second, in the case of selecting discrete objects in the world on the basis of a small two-dimensional image, context may not be very pertinent.

Regarding the second explanation, it might be that what many of the other studies were looking for was the 'meaning' of the images, especially with respect to communicative intent (for AAC purposes). For this specific use of images, and especially in a use environment where the image used is intended to be very specific to the goal, such as in a supermarket (Carmien, 2006), the fidelity of the image is much more important than its arrangement. In this case of the MAPS system it is easy to change the images to reflect the specifics of the task on a day-to-day basis (i.e., image could be changed for each instance of the prompter script that fit the user, goal and task). In any case, our results, while not conclusively supporting the conjecture that drove the experiment, suggest that further exploration of this topic would be helpful in building an effective computational task support system.

Based on our results, icons and photographic images depict real-world objects in different ways. This difference, whatever it may be, is enough to lead to errors in matching those representations to their real-world referent. The results imply that making 'generic' scripts out of icons would not be a good strategy for creating scripts to guide person with cognitive disabilities in daily tasks. Given that knowledge, the current practice of assembling prompts out of icons needs to change towards using realistic photos. For some participants and objects a generic photo may suffice, for others a generic photo may either not be available or not have rich enough detail to afford easy matching. This leads to a requirement of computationally enhanced prompting systems to support the easy use of caregiver taken photos in the creation of scripts. The MAPS system provides this flexibility in altering scripts. Specifically, there is a built-in facility for importing and using specific images for creation of scripts by computer novices, which the majority of caregivers are assumed to be. The MAPS suite of applications has been used to create complex scripts that allowed several young adults with cognitive disabilities to do complex tasks (Carmien, 2004b).

We are interested in future studies of image class and matching success, and further probing of 'best practices' of assistive technologists and rehabilitative therapists. One goal of this research, and of our future studies, is the creation of a test to individually determine the optimal match between the type of user, based on individual differences, and image attributes.

Acknowledgements

The authors would like to express our appreciation to Anja Kintsch and the Boulder Valley School District for helping to select and supply participants and space for this study, and to Gen Hudak our undergraduate apprentice for collecting the images and assisting in programming of the testing system. We would also like to thank the Hewlett Packard Company for their generous contribution of IPAC hand-held computers. This research was supported in part by grants from the Coleman Institute for Cognitive Disabilities.

References

American Psychiatric Association, Task Force on DSM-IV. (2000). *Diagnostic and statistical manual of mental disorders: Dsm-iv-tr.* Washington, DC: American Psychiatric Association. (4th, text revision. ed.).

Assistive Technology, I. (2005). Evaluware.

Beukelman, D., & Mirenda, P. (1998). *Augmentative and alternative communication* (2nd ed.). Baltimore: Paule H Brookes Publishing.

Biederman, I. (1995). Visual object recognition. In S. F. Kosslyn, & D. N. Osherson (Eds.), *An invitation to cognitive science* (2nd ed.). Boston MA: MIT Press. (vol. 2, Visual Cognition, pp. 121–165).

Braun, J. (2003). Natural scenes upset the visual applecart. *TRENDS in Cognitive Sciences, 7*(1), 7–9.

Carmien, S. (2004a). Maps website 2005, from http://l3d.cs.colorado.edu/clever/projects/maps.html.

Carmien, S. (2004b). Task support for people with cognitive impairments and their caregivers. *American Journal of Occupational Therapy, 14*(3), 1–4.

Carmien, S. (2005). End user programming and context responsiveness in handheld prompting systems for persons with cognitive disabilities and caregivers. *Paper presented at the Proceedings of CHI'05 Conference on Human Factors in Computing Systems*.

Carmien, S. (2006). *Socio-technical environments supporting distributed cognition for persons with cognitive disabilities. Unpublished Ph.D. Dissertation*. University of Colorado, Boulder, Colorado.

Davies, D., Stock, S., & Wehmeyer, M. L. (2002). Enhancing independent task performance for individuals with mental retardation through use of a handheld self-directed visual and audio prompting system. *Education and Training in Mental Retardation and Developmental Disabilities, 37*(2), 209–218.

Dixon, L. (1981). A functional analysis of photo-object matching skills of severely retarded adolescents. *Journal of Applied Behavior Analysis, 14*(4), 465–478.

Huer, M. B. (2000). Examining perceptions of graphic symbols across cultures: preliminary study of the impact of culture/ethnicity. *Augmentative and Alternative Communication,* 16: 180–185.

Mackintosh, N. J., & Bennett, E. S. (2002). IT, IQ and perceptual speed. *Personality and Individual Differences,* 32: 685–693.

Mayer-Johnson. (2001). *Introduction to the picture communication symbols*. Solana Beach, CA: Mayer-Johnson.

Mirenda, P., & Locke, P. A. (1989). A comparison of symbol transparency in nonspeaking persons with intellectual disabilities. *Journal of Speech and Hearing Disorders, 54*, 131–140.

Snell, M. E. (1987). *Systematic instruction of persons with severe handicaps*. Columbus, OH: Merrill Publishing Company.

Stephenson, J., & Linfoot, K. (1996). Pictures as communication symbols for students with severe intellectual disability. *AAC Augmentative and Alterative Communication, 12*, 224–255.

Wilkinson, K. M., & McIlvane, W. J. (2002). Considerations in teaching graphic symbols to beginning communicators. In J. Reichle, D. R. Beukelman, & J. C. Light (Eds.), *Exemplary practices of beginning communicators* (pp. 273–321). Baltimore MD: Paul H. Brookes Publishing.