Int. J. Human–Computer Studies 151 (2021) 102599

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Anachronism by Design: Understanding Young Adults' Perceptions of Computer Iconography

Abdullah X. Ali *, Erin McAweeney, Jacob O. Wobbrock

University of Washington

ARTICLE INFO

Keywords:
Elicitation study
Identification study
Icons
Participatory design
Amazon Mechanical Turk

ABSTRACT

Computer iconography in desktop operating systems and applications has evolved in style but, in many cases, not in substance for decades. For example, in many applications, a 3.5" floppy diskette icon still represents the "Save" function. But many of today's young adult computer users grew up without direct physical experience of floppy diskettes and many of the other objects that are represented by enduring legacy icons. In this article, we describe a multi-part study conducted to gain an understanding of young adults' perceptions of computer iconography, and to possibly update that iconography based on young adults' *current* mental models. To carry out this work, we gathered a set of 39 icons found on common desktop operating systems and applications and also recruited 30 young adults aged 18–22. In the first part of our study, an end-user elicitation study, we asked participants to propose sketches of icons they deemed most appropriate to trigger the functions associated with our selected icons. We elicited a total of 3,590 individual icon sketches and grouped these into a set of participant-generated icons. In the second part of our study, an end-user identification study, we showed participants the 39 icons from current operating systems and asked them to name the computing functions triggered when those icons were selected. We also asked them to identify the real-world objects, if any, those icons represented, and to tell us about their personal experiences with those objects. Finally, we conducted a second identification study with 60 new participants from Amazon's Mechanical Turk on the set of participant-generated icons we obtained from the first part of our study to see how recognizable our young adults' sketched icons were. Our study results highlight 20 anachronistic icons currently found on desktop operating systems in need of redesign. Our results also show that with increased icon production, the chances for anachronism significantly decrease, supporting the "production principle" in elicitation studies. Furthermore, our results include an updated set of icons derived from our young adult participants. This work contributes an approach to using end-user elicitation to understand users, user interface design, and specifically, icon design.

1. Introduction

The graphical representation of a computer's status and functions as icons has been prevalent since icons' inception in David Canfield Smith's *Pygmalion* (Smith, 1977), and then icons' subsequent commercial adoption in the graphical user interface of the Xerox Star (Johnson et al., 1989). The Star based its icons on everyday objects familiar to office desktops of the time. Despite the technological strides made since the advent of these early icons, amazingly, many of the same icons persist decades later, even into today's desktop computer systems. Some of these icons are even graphical representations of objects that are no longer used in most people's everyday lives. For example, the 3.5" floppy diskette—often representing the "Save" function—and the compact disk—representing the Windows "Program Manager"—are computer artifacts no longer used by most people. Today's young adult technology users have never interacted with such legacy objects, which could complicate the guessability and learnability of icons based on them, and raises interesting questions about users' mental models.

An example of the disconnect between the anachronistic objects represented by some of today's interface icons and young adults' perceptions of the objects themselves can be seen in Fig. 1, which is a tweet from a person who, holding a 3.5" floppy diskette, was told by a youth that he had "3D-printed the 'Save' icon." Rather than the physical diskette informing the meaning of the computer icon, the computer icon had informed the meaning of the physical object—it was seen as merely a plastic model of the icon. The computer icon was now the prevalent

* Corresponding author.
E-mail addresses: abdullah@axali.me (A.X. Ali), ermcaweeney@gmail.com (E. McAweeney), wobbrock@uw.edu (J.O. Wobbrock).

https://doi.org/10.1016/j.ijhcs.2021.102599
Received 6 July 2020; Received in revised form 12 January 2021; Accepted 19 January 2021
Available online 22 January 2021

Fig. 1. A tweet with an image of a physical floppy diskette. The tweet reads, "In the 'I'm getting old' department, a kid saw this and said, 'oh, you 3D-printed the 'Save' Icon.'"

object, and the physical diskette was now subject to *its* meaning.

Along with text, icons are of central importance for guessing and learning functions in any graphical user interface. Stylistic updates to icons are common, but a certain "stickiness" pervades the conceptual underpinnings of icon design. Presumably, this stickiness is so that existing users, upon receiving a software update or installing a new application, can leverage their pre-existing knowledge of icons' meanings. At the same time, newer generations of computer users are encountering more and more icons, the objects for which they have never encountered in the physical world, like the anachronistic 3.5" floppy diskette. The word "anachronistic" means belonging to a period other than that being portrayed. We operationalize the term "anachronistic icon" as computer iconography portraying physical objects that have been largely displaced by current technology. Examples of anachronistic icons besides the floppy diskette include desk calendars replaced by software applications, compact disks replaced by digital downloads, and a stopwatch replaced by a phone app. In this work, we use such icons as a use case for end-user elicitation and identification studies with the goal of understanding young adult users' familiarity with the objects depicted in these icons and gain insight into what alternative icons for functions associated with these icons would look like, according to our young adult participants.

To understand young adults' perceptions of the objects represented by the icons in current desktop operating systems and applications, we assembled a set of 39 icons found on Windows 10 and Mac OS X that feature plausibly anachronistic objects. Based on the set of 39 icons and the functions they are associated with, we sought to answer the following research questions:

- What icons would young adults sketch to trigger each one of the 39 computer functions?
- How familiar are young adults with the objects represented by the 39 selected icons?
- How identifiable is a set of icons elicited from young adults?

To answer our questions, we conducted a multi-part study. First, we recruited 30 young adult technology users aged 18–22 to participate in a two-part study. The first part was an icon elicitation study based on Wobbrock *et al.*'s (Wobbrock et al., 2005, Wobbrock et al., 2009) end-user elicitation method. In this study, we presented our participants with 39 descriptions of computing functions that currently have an icon representing a plausibly anachronistic object (*e.g.*, "save"). We asked our participants to sketch (and describe) icons of their own making that would trigger these functions. We elicited a total of 3,590 icons from our 30 participants, or an average of about three icons per function from each participant. We then clustered similar icons to arrive at a set of 39 participant-generated icons.

In the second part of our two-part study, we conducted an identification study with our 30 participants based on Ali *et al.*'s (Ali et al., 2019) end-user identification method. In this part, we presented our participants with cards, each showing one plausibly anachronistic icon, and asked them to identify the computing function that the icon would trigger when selected. We also asked our participants how familiar they were with the real-world objects depicted by the icons. Finally, to assess the set of participant-generated icons that we derived from our elicitation study, we conducted a second identification study with 60 participants recruited online from Amazon's Mechanical Turk using Ali *et al.*'s Crowdlicit system (Ali et al., 2019).

We conducted open-coding (Strauss and Corbin, 1997) on all 3,590 icons that we collected in our elicitation study and formulated a taxonomy of computing iconography. We found that almost half of our 3,590 participant-generated icons were of new concepts, while the other half were drawings of existing anachronistic icons. When assembling the participant-generated set of icons from our elicitation study, we found that 17 out the 39 icons remained anachronistic, and 22 icons were of new concepts. We also found that 73% of all elicited icons were representations of physical objects. The remaining 27% were made up mostly of text and abstract shapes. We also found that the propensity for elicited icons to be anachronistic decreased significantly from the first elicited icon to subsequent icons, confirming that the "production principle" (Morris et al., 2014) does, indeed, increase the novelty of elicited symbols, at least in this context.

In our laboratory-based identification study, we found that there were only 16 plausibly anachronistic physical objects that all of our 30 participants had ever used. Furthermore, our identification study with young-adult participants resulted in the correct identification of the functions triggered by 31 of the 39 plausibly anachronistic icons. Of the eight icons that had their functions identified incorrectly by our young adult participants, five of them had new concept icons from the elicitation study. These five new concept icons did indeed improve identifiability, as they were identified correctly by our second identification study—an online study with 60 new participants from Amazon's Mechanical Turk. In this second identification study, the 60 online participants were able to correctly identify 34 of the 39 user-generated icons from our elicitation study. Three of the five incorrectly identified icons were new concept icons, and two were anachronistic icons.

Finally, as a result of all three parts of our multi-part study (in-lab elicitation and identification, and an online identification study), we derived a set of 39 participant-generated icons that included three icon types: *Anachronistic by Elicitation, Anachronistic by Identification*, and *New Concept* icons. This set of icons contained 20 new concept icons and 19 anachronistic icons: 15 *Anachronistic by Elicitation* and four *Anachronistic by Identification.*

This article contributes empirical results of an elicitation study, two identification studies, a taxonomy of computer iconography, and a set of participant-generated icons based on the results of the three studies. It also provides empirical support for the "production principle" (Morris et al., 2014) for generating novel designs. Generally, this work can be of use to researchers understanding young adult users, user interface designers, and specifically, iconographers looking to design the next generation of icons for tomorrow's graphical user interfaces.

2. Related work

Prior work related to the current research includes (1) computer iconography research in human-computer interaction (HCI), and (2) end-user elicitation studies. These topics are each addressed in turn.

2.1. Computer iconography research in HCI

The prevalent use of icons in today's highly graphical user interfaces is of central importance, as icons can communicate status and function meanings faster and more effectively than text (Collins and Lerner, 1982, Ware, 2010). Icons first appeared in David Canfield Smith's 1977 *Pygmalion* system for programmers (Smith, 1977). Smith's icons combined visual depictions with behavior; in this case, the execution of computer programs. Later, Smith joined Xerox PARC and the team working on the Xerox Star. The Star employed Smith's icons, reworked to represent office concepts rather than programming concepts. In this way, the desktop metaphor was born, along with the original set of icons designed to convey to knowledge workers about the Xerox Star's status, features, and functions, making learning and operating the Star more intuitive. According to Smith, the Star's icons were designed to be "visible, concrete embodiments of the corresponding physical objects" (Smith, 1982). That is, the direct association between the real-world physical object and its computer-icon counterpart was considered a deliberate, even vital, one.

But today, this association is no longer maintained for many plausibly anachronistic icons, especially for young-adult users who have never encountered many icons' real-world physical counterparts. Young-adult users who have never relied on a pushbutton calculator or seen a 3.5" floppy diskette in person do not have the same familiarity with objects represented by certain icons found in today's operating systems and applications. Young-adult users of today, unfamiliar with these anachronistic objects, might not draw the same intended associations as now older users once did.

Yan (Yan, 2011) outlined an icon classification system in interface design by drawing on theories of cognitive psychology and semiotics. Yan's classification had four categories to describe computer iconography: icon type with basic image feature, abstract signifying icon type, text icon type, and number icon type. All the plausibly anachronistic icons that are the subject of investigation in this work fall in the first category of "icons with basic image feature," as they represent images of anachronistic objects.

Ho and Hou (2015) examined mobile app icons and their relationships to users. Ho and Hou established eight factors of attractiveness for designing app icons: artifact imitation, cartoon elements, 3-D effect, color, brilliant logo, dynamic elements, appropriate function, and novelty. App designers can look to these elements to make their app icons more attractive to potential users in app stores.

Mavri et al. (2016) conducted a user study to develop icons to support the tasks of information seekers in academic document triage interfaces. In their study, Mavri *et al.* asked participants to draw how they visualized certain elements of a of an academic document (*e.g.,* author, abstract.) They found that elements of an academic document can be represented in an icon using spatial properties, formatting (such as line thickness), linguistic properties (like including an alphanumeric character in the icon), and homonym properties (representation of real-world objects). The icons can be representative of archetypal forms to signify their meanings, like the use of a face with a mustache and glasses to represent an author, or a test tube to represent a science application. A similar approach was utilized by McKnight and Read (McKnight and Read, 2009) to design the "record" button icon for a mobile app by eliciting icon ideas from children (aged 8–10). McKnight and Read found it challenging to design an easily- recognizable icon for "record." Their work further gives evidence of the need for well designed and easily recognizable icons. Well designed icons enhance usability as shown by Chen et al. (2013), who studied the iconography of social networking sites and found that well designed icons illustrate higher correctness and familiarity, coupled with less complexity.

Some researchers working on icon creation based their approach on ethnographic interviews with professional icon designers to automatically generate icons, like the work of Zhao et al. (2020).

In contrast to prior work on iconography, the current work investigates the substance of icons that represent plausibly anachronistic objects prevalent in desktop operating systems and applications. We contribute a study of young adults' perceptions of such icons and insights into what replacement icons these young adults would envision for the plausibly anachronistic icons.

2.2. End-user elicitation studies

The practice of employing users to propose interaction designs is a popular design method in the field of human-computer interaction (HCI). Perhaps the earliest example dates back to 1984 when Good et al. (1984) had users propose command terms to design an intuitive command-line interface. Wobbrock et al. (2005), Wobbrock et al. (2009) formalized a similar approach, initially around gestural interactions, adding conflict resolution techniques and a formula for agreement calculation. Many have since replicated Wobbrock *et al.*'s methodology in HCI research, with now nearly 300 published studies employing the method. The methodology has been widely used to explore user-defined gestures to interact with a broad variety of technologies (Connell et al., 2013, Freeman et al., 2013, Gheran et al., 2018, Mauney et al., 2010, Modanwal and Sarawadekar, 2018, Obaid et al., 2012, Piumsomboon et al., 2021, Tan et al., 2017). Obaid et al. (2012) used the method to elicit full-body gestures for controlling humanoid robots. Piumsomboon et al. (2013) used it to capture user-defined interactions for augmented reality. Tan et al. (2017) used the method to elicit micro hand gestures as input for cycling.

The end-user elicitation method has been used to elicit other input modalities. For example, Morris et al. (2012) used it to elicit speech and gesture interactions for TV-based web browsing. Nebeling et al. (2014) and Ali et al. (2018) used the method to elicit voice commands. McAweeney et al. (2018) used the method to elicit graphical representations of gestures to create user-driven design principles for gesture representation.

Since Wobbrock *et al.'s* (Wobbrock et al., 2005, Wobbrock et al., 2009) formalization of the method, there has been work published extending the method itself. Several versions of the agreement equation have been proposed (Findlater et al., 2012, Vatavu and Wobbrock, 2015, Vatavu and Wobbrock, 2016) and some have raised arguments questioning their validity (Tsandilas, 2018). Morris et al. (2014) proposed techniques to enhance participants' creativity, which we employ in our work. Due to the elicitation method's popularity, Ali *et al.* created a platform (Ali et al., 2019) to allow for conducting elicitation studies with online participants on a global scale. Their platform also enhances the efficiency of analyzing elicitation studies by using online crowds and machine learning (Ali et al., 2018).

To the best of our knowledge, the current work is the first to utilize the end-user elicitation method to elicit computer icon designs from end-users.

3. Anachronistic icons

We scoured current desktop operating systems and applications—both Windows 10 and Mac OS X—and assembled a set of 39 plausibly anachronistic icons that represented 38[1] real-world physical objects no longer as widely used as they perhaps once were. (We call these icons "plausibly anachronistic," because for some young-adult users, their physical objects might still be conceivably familiar.) One of the authors collected icons that depicted physical object found on desktop operating systems, and then all three authors discussed whether or not to include each of these icons in the list of plausibly anachronistic icons.

We reiterate our definition of anachronistic icons to be computer

[1] Functions "#2. searches" and "#3. zooms" shared the magnifying glass object.

iconography portraying physical objects that have been largely displaced by current technology. Fig. 2 displays 39 icons and their associated functions. A prevalent example of a plausibly anachronistic icon is the 3.5" floppy diskette (icon #20), the physical version of which is rarely still used except on antiquated computers, but whose semblance still pervades user interfaces as the "save" icon. Other possibly outmoded real-world objects for young adults might include the magnifying glass (#2, #3), "snail mail" (#6, #26), print photographs (#8), printed books (#13, #14), paper calendars (#17, #18), analog clocks and watches (#17, #24, #38), compact discs (#31), filament light bulbs (#34), and analog magnetic compasses (#39), among others. Table 1 displays the source system and original system of each icon.

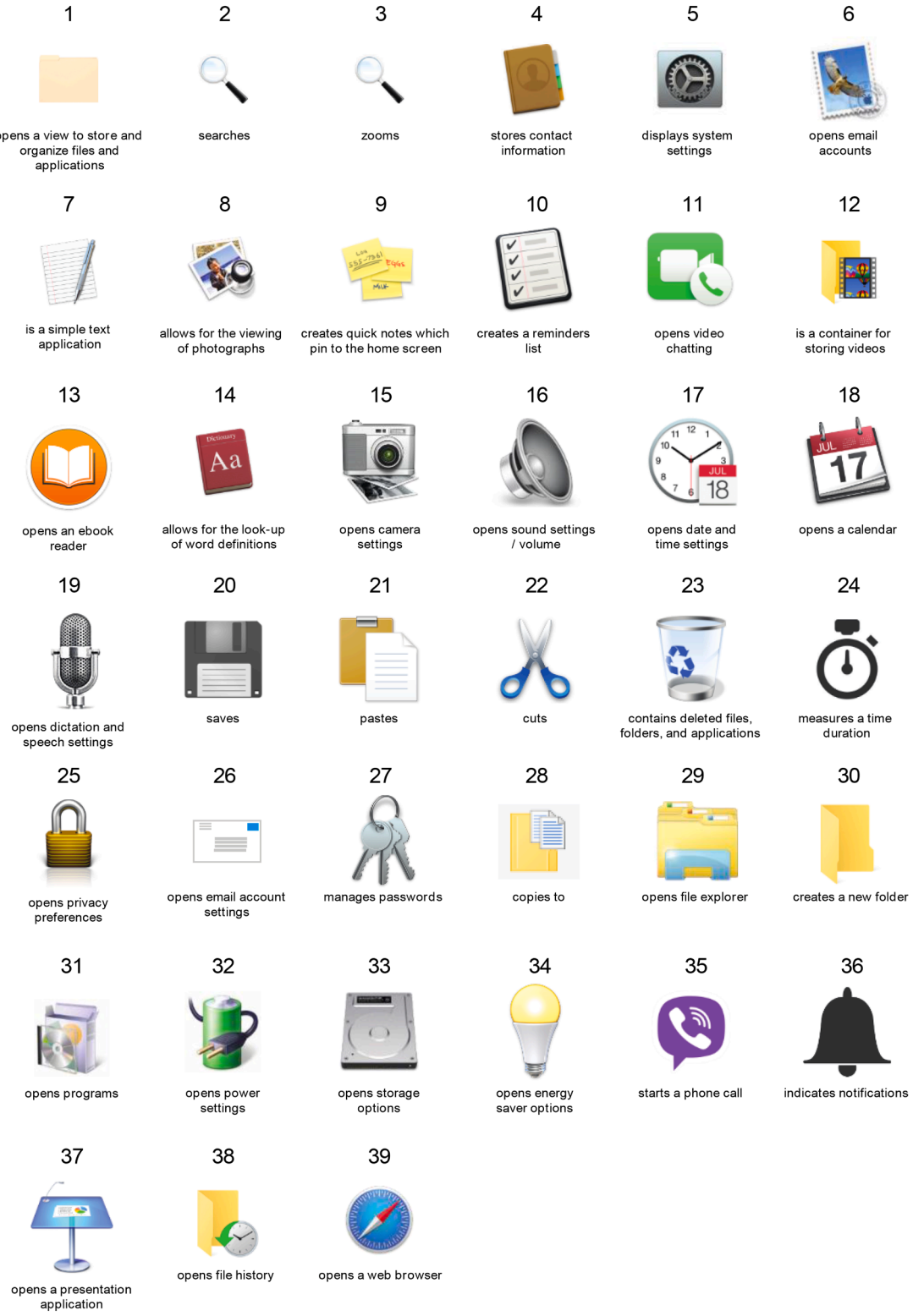

Fig. 2. The 38 plausibly anachronistic icons and the 39 functions they trigger. (Note that icons 2 and 3 are the same, but for different functions.)

Table 1
The 39 computer functions of our plausibly anachronistic icons. Also shown are the systems from which our icons were taken and the systems in which they first appeared.

#	Function ("An icon that …")	Source	Origin, Year first appeared
1	...opens a view to store and organize documents, files, and applications	Windows 10	Xerox Star, 1981
2	...searches	Windows 7	Windows 95, 1995
3	…zooms	Windows 7	Windows 95, 1995
4	...stores contact information	Mac OS X 10.11.6	Windows 1, 1985
5	...displays system settings	Windows 10	Windows 95, 1995
6	...opens email accounts	Mac OS X 10.11.6	Mac OS X, 2001
7	...is a simple text application	Mac OS X 10.11.6	Xerox Star, 1981
8	...allows for the viewing of photographs	Mac OS X 10.11.6	NeXTSTEP, 1989
9	...creates quick notes which pin to the home screen	Mac OS X 10.11.6	Mac OS 7, 2013
10	...creates a reminders list	Mac OS X 10.11.6	OS X 10.8 "Mountain Lion", 2012
11	...opens video chatting	Mac OS X	Mac OS X 10.6, 2010
12	...is a container for storing videos	Windows 10	Windows 98, 1998
13	...opens an ebook reader	Mac OS X 10.11.6	Mac iOS 4, 2010
14	...allows for the look-up of word definitions	Mac OS X 10.11.6	OpenStep, 1994
15	...opens camera settings	Mac OS X 10.11.6	Mc OS X 10, 2001
16	...opens sound settings / volume	Mac OS X 10.11.6/ Windows 10	Xerox Star, 1981
17	...opens date & time settings	Mac OS X 10.11.6	Apple Lisa, 1983
18	...opens a calendar	Mac OS X 10.11.6	Windows 1 1985
19	...opens dictation and speech settings	Windows 10	Xerox Star, 1981
20	...saves	Mac OS X 10.11.6	Xerox Star, 1981
21	...pastes	MacOS X 10.11.6	Apple Lisa, 1983
22	...cuts	MacOs X 10.11.6	Xerox Star, 1981
23	...contains deleted files, folders, and applications	Windows 7	Xerox Star, 1981
24	...measures a time duration	Windows 8	Apple Lisa, 1983
25	...opens privacy preferences	Windows 7	Macintosh System 7, 1991
26	...opens email account settings	Windows 10	Windows 3.1, 1992
27	...manages passwords	Windows 10/ Mac OS X 10.11.6	Windows 3.1, 1992
28	...copies to	Windows 10	Xerox Star, 1981
29	...opens file explorer	Windows 7	Xerox Star, 1981
30	...creates a new folder	Windows 10	Xerox Star, 1981
31	...opens programs	Windows 10	Windows 3, 1995
32	...opens power settings	Windows 10	Windows NT 3.1, 1992
33	...opens storage options	Windows 10	Xerox Star, 1981
34	...opens energy saver options	Mac OS X 10.11.6	Mac OS X, 2001
35	...starts a phone call	Google Gmail	Gmail, 2009
36	...indicates notifications	Google	Google, 2003
37	...opens a presentation application	Mac OS X 10.11.6	NeXTSTEP, 1989
38	...opens file history	Windows 10	Xerox Star, 1981
39	...opens a web browser	Mac OS X 10.11.6	Mac OS X Panther, 2003

4. Understanding young adults' perceptions of icons

We conducted a two-part study with 30 young-adult technology users to understand what icons they would create to trigger functions that are currently triggered by our plausibly anachronistic icons. We also sought to understand young adults' perceptions of anachronistic icons and the experiences they have of the real-world objects those icons portray. Our study also resulted in a set of user-generated icons that we validated by conducting a crowdsourced identification study with 60 new participants to see how recognizable the new user-generated icons were to general computer users.

4.1. Young-adult participants

We recruited 30 young adults to participate in our two-part study using flyers on and around our university campus. At the time of our study, our 18–22 year-old participants were born between the years 1994–1998. All participants were students at our university majoring in a wide range of fields including computer science, bioengineering, informatics, political science, and humanities. About half (57%) of our participants were from the United States, 23% were from China, and others were from numerous other countries such as Switzerland and Nepal. Fifty-seven percent of our participants were women. Most of the participants (87%) did not have any professional design background. About half of the participants reported spending 6–10 hours a day using a computer, 3–5 hours using a mobile device, and about 70% of them spent less than 2 hours a day using physical objects (*i.e.,* books, wrenches). Table 2 reports additional demographic information for our participants.

4.2. New icon elicitation study

An end-user elicitation study is a user-centered design methodology in which end-users are presented with the effect of an input, known as a *referent*, and are asked to propose the input itself, known as a *symbol*, that would invoke that referent (Wobbrock et al., 2005, Wobbrock et al., 2009). In Wobbrock *et al.'s* original work (Wobbrock et al., 2005, Wobbrock et al., 2009), symbols were stylus or hand gestures. In work by Morris et al. (2012), Nebeling et al. (2014), and Ali et al. (2018), symbols were voice commands. In the current study, symbols are sketches of computer icons that would trigger computer functions, which are the referents. For example, an icon of a globe, when triggered, could open a Web browser. Similarly, an icon of a world map could do the same. The globe and the map are symbols, and opening a Web browser is the referent.

In an attempt to reduce "legacy bias" (Morris et al., 2014), where participants propose only familiar symbols from systems they already know, we conducted the elicitation study first before the identification study, thus limiting exposing our participants to plausibly anachronistic icons in the identification study.

The elicitation study session took about two hours to complete. In a session, participants were given a card that displayed text descriptions of computer functions (*e.g.*, "open a calendar") and were asked to sketch as many icons as they could reasonably devise to trigger each function. Having participants propose multiple icons, rather than just one, for each referent in an elicitation study is a legacy bias reduction technique known as the "production principle" set forth by Morris et al. (2014). Although the production principle has been proposed previously, little empirical evidence exists as to its effectiveness. As our results show, the more icons participants proposed, the less likely they were to be anachronistic in nature.

We asked our participants to rate each of their icon sketches on how well they felt each icon matched its intended function, prompted by the following: *"The icon I drew is a good match for its intended purpose."* Likert-type ratings were on a scale from 1–7, with 1 being "strongly disagree" and 7 being "strongly agree." We also asked our participants to rate the familiarity of the icon they drew on a 1–3 scale as such: 1 was "I am not familiar. I have never seen it before;" 2 was "I am somewhat familiar. I am not sure where I have seen it before;" and 3 was "I am very familiar. I know where I have seen it before."

4.2. Anachronistic icons identification study

In the "next steps" section of their popular paper on end-user gesture elicitation, Wobbrock et al. (2009) (p. 1091) describe a validation procedure whereby elicited gestures would be shown to new participants who guess what each gesture does. Whereas elicitation studies show referents and request symbols, such a validation procedure is the

Table 2
Demographic information for our 30 young-adult participants.

Demographic		N = 30	Demographic		N = 30
Gender	Men	13 (43%)	First device owned	Desktop	13 (43%)
	Women	17 (57%)		Laptop	3 (10%)
	Non-binary	0		Mobile phone	6 (20%)
Country of origin	USA	17 (57%)		Other	8 (27%)
	China	7 (23%)	Total hours per day using a computer (self-reported)	< 1 hour	0
	Switzerland	1 (3%)		1–2 hours	7 (23%)
	South Korea	1 (3%)		3–5 hours	7 (23%)
	Nepal	1 (3%)		6–10 hours	14 (47%)
	Indonesia	1 (3%)		> 10 hours	2 (7%)
	India	1 (3%)	Total hours per day using a mobile device (self-reported)	< 1 hour	1 (3%)
	Bangladesh	1 (3%)		1–2 hours	9 (30%)
Age	18	8 (27%)		3–5 hours	17 (57%)
	19	8 (27%)		6–10 hours	3 (10%)
	20	4 (13%)		> 10 hours	0
	21	7 (23%)	Total hours per day using a physical object (books, paint brushes, wrenches, etc.)	< 1 hour	9 (30%)
	22	3 (10%)	(self-reported)	1–2 hours	12 (40%)
Any professional design background?	Yes	4 (13%)		3–5 hours	7 (23%)
	No	26 (87%)		6–10 hours	2 (7%)
Age started using computers	3–5	8 (27%)		> 10 hours	0
	6–8	16 (53%)	Preferred desktop operating system	Mac OS	16 (53%)
	9–11	5 (17%)		Windows	14 (47%)
	12–16	1 (3%)	Preferred mobile operating system	Apple iOS	20 (67%)
	17+	0		Google Android	9 (30%)
First operating system	Mac OS	3 (10%)		Blackberry OS	1 (3%)
	Windows	26 (87%)			
	MS-DOS	1 (3%)			

reverse, showing symbols and requesting referents. Ali et al. (2019) formalized this procedure, terming it an "end-user identification study" as the complement to an "end-user elicitation study" and used it to validate the results of their elicitation study designing gestures and voice commands to interact with a Web browser. Ali et al. (2019) used the end-user identification method to find referent agreement as a measure of consensus among participants indicating which referent should be triggered by a given symbol. They measured the referent accuracy by comparing the referent with the highest consensus to the original referent for the symbol to determine if they were a match.

In an end-user identification study, symbols could be stroke-gesture sketches, freehand gesture videos, voice commands, bitmapped or sketched icons, or even auditory feedback ("beeps," "bonks," and "dings"). Upon being presented with each symbol, a participant hazards a guess at what that symbol does or means. For example, an identification study of auditory feedback might ask participants to guess the current state of the computer when a certain sound is played. Generally, participants propose referents without complete knowledge of the possible referents in a system. Researchers conducting the study aggregate participants' proposed referents based on referents' similarity to find the referent for each symbol with the highest consensus. The result is an indication of how well the symbol conveys its intended referent.

We asked our 30 young-adult participants to come back at a later date—after completing the elicitation part of the study—to complete an identification study of the icons in Fig. 2 in order to receive their total compensation of $50 for participation in both parts of the study.

In the second part of the study, we handed the participants a deck of cards. The cards were numbered, and each card had a single plausibly anachronistic icon on it from Fig. 2. Upon viewing a card, participants answered questions regarding the icon depicted. In addition to asking what function the icon on the card triggered when clicked, we asked participants to identify what real-world object was depicted by the icon, and whether participants had ever used the object itself, and when. Thus, we gathered data making it possible to examine whether there was any relationship between young adults' abilities to identify what an icon *does*, and what the icon's real-world object *is*, for our plausibly anachronistic icons.

4.3. Young-adult-generated icons identification study

The first part of our study, the end-user elicitation study, resulted in a set of user-generated icons. To assess this set of new icons, we conducted a second end-user identification study online using Ali *et al.'s Crowdlicit* system (Ali et al., 2019) (Fig. 3). We recruited 60 new participants from Amazon's Mechanical Turk (MTurk.) Participants who accepted the human intelligence task (HIT) on MTurk were directed to a unique URL created by the Crowdlicit system in which they participated in a 15-minute study. In the study, participants assumed a total of 39 tasks, where each task showed one user-generated icon in isolation with no further context. Upon viewing an icon (the "symbol"), participants were prompted to name the computing function that would be triggered by clicking that icon (the "referent"). Upon completing all 39 tasks, participants received a code that they entered back in the MTurk portal to indicate their completion of the study and ensure their payment of $3.75. (We based our payment on our city's minimum wage of $15/hour.) Participants also filled out a demographics survey upon completing the 39 tasks.

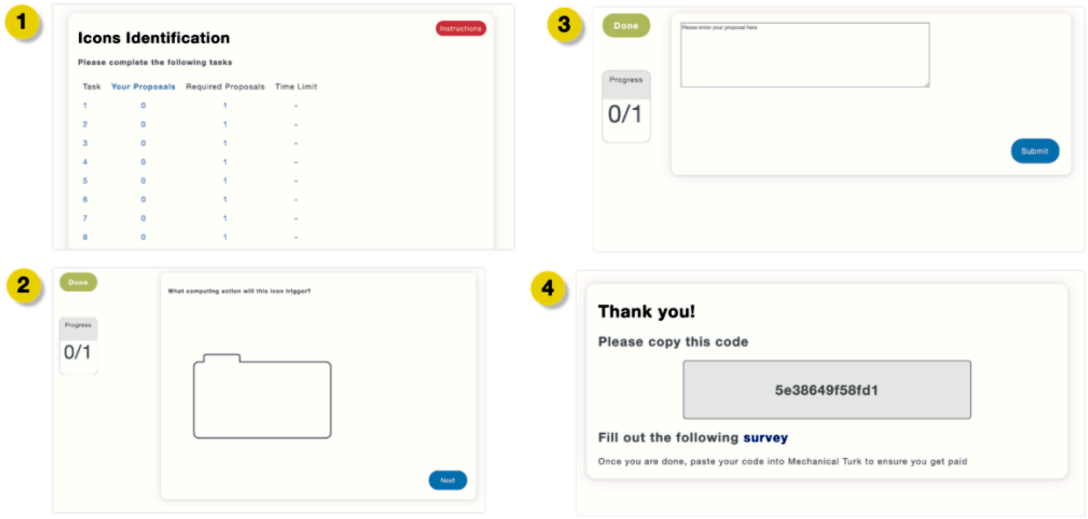

Fig. 3. The Crowdlicit interface. (1) The task list of 39 user-generated icons to identify. (2) The prompt "what computing action will this icon trigger?" and a basic image of the icon. (3) The interface to identify the referent triggered by the icon. (4) A thank you page with unique completion code and a link to the demographics survey.

Of the 60 new participants who completed the HIT, 42 filled out our demographics survey. Table 3 shows the demographics of these 42 MTurk participants. More than half of the participants (64%) were men. The majority of the participants (69%) were from the United States; 21% were from India, and the rest were from other countries including Italy and Canada. The average age of our participants was 31.10 (*SD*=7.97) years. Most of our participants (72%) indicated that they spend more than 6 hours a day using a computer. Seventy-nine percent of the participants spend between 1–5 hours per day using a mobile device. Seventy-two percent of the participants spend 2 hours or less using

Table 3
Demographic information for our 42 participants recruited from Amazon's Mechanical Turk.

Demographic		N = 42	Demographic		N = 42
Gender	Men	27 (64%)	Total hours a day using a computer (self-reported)	< 1 hour	1 (2%)
	Women	15 (36%)		1–2 hours	2 (5%)
	Non-binary	0		3–5 hours	9 (21%)
Country of origin	USA	29 (69%)		6–10 hours	18 (43%)
	India	9 (21%)		> 10 hours	12 (29%)
	Italy	1 (2%)	Total hours a day using a mobile device (self-reported)	< 1 hour	4 (10%)
	Canada	1 (2%)		1–2 hours	21 (50%)
	Pakistan	1 (2%)		3–5 hours	12 (29%)
	Ireland	1 (2%)		6–10 hours	2 (5%)
Average age	31.10 (*SD*=7.97) years		> 10 hours		3 (7%)
Professional design background?	Yes	11 (26%)	Total hours a day using a physical object (books, paint brushes, wrenches, etc.) (self-reported)	< 1 hour	10 (24%)
	No	31 (74%)		1–2 hours	21 (50%)
Age started using computers	3–5	10 (24%)		3–5 hours	7 (17%)
	6–8	13 (31%)		6–10 hours	2 (5%)
	9–11	11 (26%)		> 10 hours	2 (5%)
	12–16	2 (5%)	Preferred desktop operating system	Mac OS	8 (19%)
	17+	6 (14%)		Windows	33 (79%)
First operating system	Mac OS	4 (10%)		Other	1 (2%)
	Windows	35 (83%)	Preferred mobile operating system	Apple iOS	17 (40%)
	Other	3 (7%)			
First device owned	Desktop	35 (83%)		Android	25 (60%)
	Laptop	6 (14%)			
	Other	1 (2%)			

physical objects.

5. Results: Icons and perceptions

We present the results of our studies in the following subsections, including the icons proposed by our 30 young-adult participants in the elicitation study, participants' perceptions of plausibly anachronistic icons (Fig. 2), and participants' experiences, if any, of the real-world objects the icons represent. We also present the results of an additional crowdsourced identification study we conducted with 60 new Mechanical Turk participants on the set of user-generated icons that resulted from our in-lab elicitation study with our young-adult participants. We take each of these in turn.

5.1. Eliciting user-generated icons

Our 30 young-adult participants offered a total 3,590 icon sketches for 39 referents (Table 1). Participants were encouraged to propose as many icon sketches as they wanted for each referent. On average, 3.07 (SD=0.68) icons were sketched per referent by each participant (Fig. 4). Overall, then, the average number of icon sketches elicited per referent was 91.92 (SD=3.83). We conducted open-coding analysis (Strauss and Corbin, 1997) on the entire set of icons, and then followed Wobbrock *et al.'s* (Wobbrock et al., 2009) method of similarity-based clustering to derive a set of 39 user-generated icons.

5.1.1. Understanding the set of elicited icons

One of the authors of this article took an open-coding approach (Strauss and Corbin, 1997) to analyze the entire set of 3,590 icon sketches elicited for 39 referents from our 30 participants. After discussing the codes iteratively with the other authors, we generated a set of 14 codes to describe the icon sketches we collected. The codes described whether an icon sketch represented a physical object, furthered the desktop metaphor, or contained a metaphor from newer technologies such as mobile devices and their interfaces. The codes also included anachronisms of both objects and practices, representations of computer hardware and actions, representations of natural elements, and physical activities or body parts. Other codes described shapes present in sketches like abstract squares or lines. The codes included design conventions—e.g., "hamburger menu." Table 4 shows the code book we formulated and used to analyze the icons, as well as the number of icons and percentage classified for each code, and an illustrative example of the code.

We found that the majority of icons (73%) proposed by the young-adult participants were representations of physical objects. In addition, 60% of all icons extended the desktop metaphor, and 55% of all sketches were of existing plausibly anachronistic icons. The sketches were also influenced by new technologies, as 20% of icons had metaphors based on mobile platforms and others such as "the cloud." Thirty percent of all icons were of abstract shapes such as arrows and rectangles. Finally, 20% of all icons had alphanumeric text in them.

5.1.2. Creating a set of user-driven icons

Following Wobbrock *et al.'s* method (Wobbrock et al., 2005, Wobbrock et al., 2009), for each of our 39 referents, we grouped participants' icon sketches based on their similarity. We took the sketches with the highest consensus for each referent and illustrated them as clean representative line drawings as shown in Fig. 5. The figure shows three example sketches provided by three separate participants (P007, P020, and P027). The three sketches were in response to referent 21 *"draw an icon that pastes."* These examples were of a glue bottle, which was the icon concept of the highest consensus among our participants, which led us to select it as the new icon for paste.

Following the same approach, we produced a set of 39 user-generated icons (Fig. 6). We list the referents associated with each icon in Table 5, distinguish whether the user-generated icon is of a new concept or of the same plausibly anachronistic icon, and report the median "match" and "novelty" scores. (The match score is a user-reported score on a 7-point Likert scale assessing the proposed sketch on how well it matches the referent. The novelty score is a 3-point scale assessing the novelty of the concept that the sketch depicts.)

When assembling the set of user-generated icons, we omitted the first elicited icon for each referent by each participant because participants tended to propose legacy interactions first (Morris et al., 2014)—in other words, their first impulses were to simply provide an icon with which they were already familiar. Because we did not show participants the current—plausibly anachronistic—icon being used for each referent and did not tell them they were not allowed to draw this icon, 70% of the first-elicited icons were plausibly anachronistic. The second elicited icons, comparatively, contained 56% anachronistic icons. In general, with more icons proposed for each referent, the chances that the icon was anachronistic, compared to the first icon, went down significantly (Fig. 7). A mixed logistic regression model (Stiratelli et al., 1984; Gilmour et al., 1985) for the chances of anachronism by icon production order shows a significant main effect ($\chi^2(5, N=3{,}590) = 65.87$, $p<.0001$). *Post hoc* pairwise comparisons corrected with Holm's sequential Bonferroni procedure (Holm, 1979) show that the first icon was significantly more likely to be anachronistic than icons proposed second, third, fourth, or fifth, but icons in these positions were not significantly different. (Icons proposed in the sixth position were not significantly less likely to be anachronistic than the first icons proposed, but this result is unreliable due to having only N=19 icons out of 3,590 proposed sixth.)

When calculating icon agreement rates for each of the 39 referents,

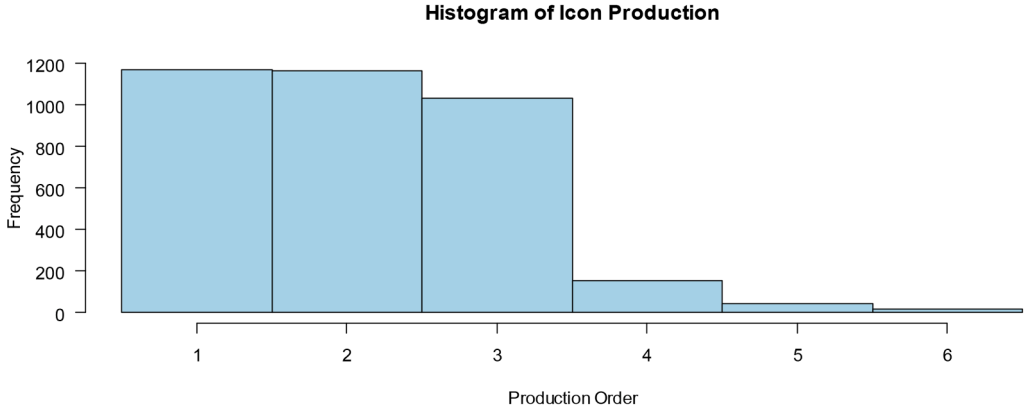

Fig. 4. The total number of icons proposed by 30 participants in our study by production order (1st, 2nd, 3rd, etc.). All participants proposed at least two icons for each computing function, but very few proposed as many as five or six icons for a computing function

Table 4
The codebook and the breakdown of our coding of the 3,590 elicited icons.

Code	Description	N = 3,590	Example
Object	Representation of a physical object for its literal meaning	2,612 (73%)	Book
Metaphor	Representation of a real-world concept or practice for an associated meaning		
Desktop	Furthering the desktop metaphor	2,157 (60%)	Use of file folders
New Technology	Use of new technology metaphors	574 (16%)	The cloud
Mobile	Furthers metaphors from mobile	150 (4%)	Home button
Anachronistic Practice	Something we used to do	56 (2%)	Store contacts in a physical address book
Computer Hardware	A non-anachronistic object used for computer function	162 (5%)	USB Drive
Computer Action	Something that happens on the computer screen	17 (0.4%)	Typing, clicking
Rare	An object that is rarely used	200 (6%)	Microscope; not a common household object like a knife or notebook
Body Part	A part of the human body	309 (9%)	An eye
Physical Activity	Represents physical activity	39 (1%)	Hand motion
Text	Text including alphanumerical characters	729 (20%)	The word "save"
Shape			
Abstract	General shape	1,088 (30%)	Circle, rectangle, arrow
Design Convention	Familiar design convention	365 (10%)	Hamburger menu, 3 dots for typing
Nature	Natural elements	171 (5%)	Animals, plants, and other natural elements
Brand			
Brand Name	Text use of a brand name	26 (0.7%)	"Skype", "Gmail"
Logo	Use of image representation of a brand	137 (4%)	"Apple logo"
Brand Influence	A visual representation of an element influenced by a brand practice	73 (2%)	Apple's use of groups of apps
Meta GUI	Use of existing UI elements in creating new icons	332 (9%)	Dropdown menu item; mouse cursor
Anachronistic object	The use of an anachronistic object		
Existing Icon	Use of an existing anachronistic icon	1,965 (55%)	A floppy disk
New Anachronism	Use of a new anachronistic icon	237 (7%)	Binoculars

we did not use Wobbrock *et al.'s* original agreement formula (Wobbrock et al., 2005, Wobbrock et al., 2009), because we had an unequal number of icons elicited for each referent; instead we used Morris's max-consensus score (Meredith Ringel, 2012). The max-consensus score refers to the percentage of participants who proposed the most frequent icon. The scores range from [0–1], with 1.00 meaning that all the participants proposed the same icon and were in complete agreement.

Fig. 8 shows the max-consensus rates for all 39 user-generated icons. On average, a quarter of our participants agreed on what the icon to trigger a function should be. The mean max-consensus score was 0.25 (SD=0.09). The icon for referent #22—the function "cut"—had the highest max-consensus score at 0.52, *i.e.,* half the participants proposed the same icon (a pair of scissors), which we included as a plausibly anachronistic icon because of its basis in the desktop metaphor from an era when physical paper was more prevalent on physical desktops. The icon for referent #31—"open program"—had the lowest score of 0.07. The icon for referent #31 is merely the word "Launch."

5.2. Identifying anachronistic icons

We asked our participants to identify the objects depicted by each one of our plausibly anachronistic icons (Fig. 2), whether they have used the depicted real-world object and when, and what computer feature or function would be triggered if the icon were selected. Following Ali *et al.'s* (Ali et al., 2019) end-user identification method, we grouped the proposed referents based on their similarity, calculated referent agreement rates, and referent accuracy.

5.2.1. Referent agreement

We used the referent agreement formula that Ali et al. (2019) based on Wobbrock *et al.'s* (Wobbrock et al., 2005, Wobbrock et al., 2009) symbol agreement formula:

$$A_s = \sum_{P_i \subseteq P_s} \left(\frac{|P_i|}{|P_s|} \right)^2 \quad (1)$$

In Eq. 1, A_s is the agreement of referents proposed for symbol *s,* P_s is the set of all referents proposed for symbol *s*, and P_i is a subset of similar referents in P_s. Fig. 9 lists the agreement scores for all 39 anachronistic icons.

There were six icons with perfect agreement—*i.e.,* all 30 young-adult participants proposed the same referent for the icon. Despite getting a perfect identification score of 1.00, two of these icons exhibited new icon design concepts from our elicitation study (Fig. 6). One was referent #8, "viewing photographs," and the other was referent #11, "making a video call." The eight icons with above 0.90 referent agreement, or identification scores, are listed in Table 6 along with the referent identified by the participants and the user-generated icon for each referent that resulted from our elicitation study.

There were seven icons with agreement scores lower than 0.30. They were: #19. "speech and dictation settings;" #21. "paste;" #25. "privacy settings;" #28. "copies to;" #30. "create a new folder;" #31. "open program;" and #38. "file history." Six of these seven icons employed new icon concepts in our user-generated icon set. Table 7 lists these icons, the

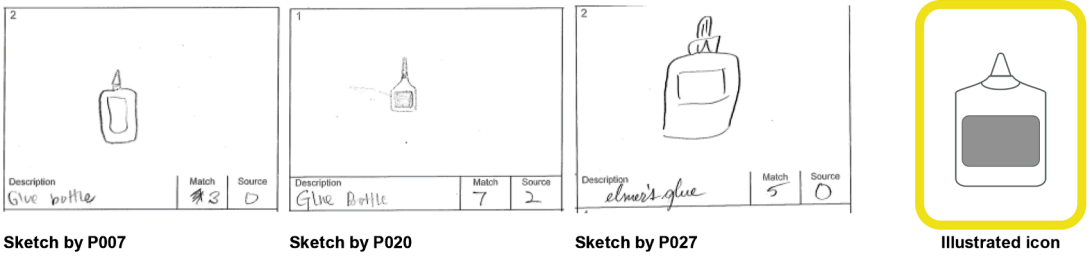

Fig. 5. An example of the illustrated icon we created based on participants' sketches.

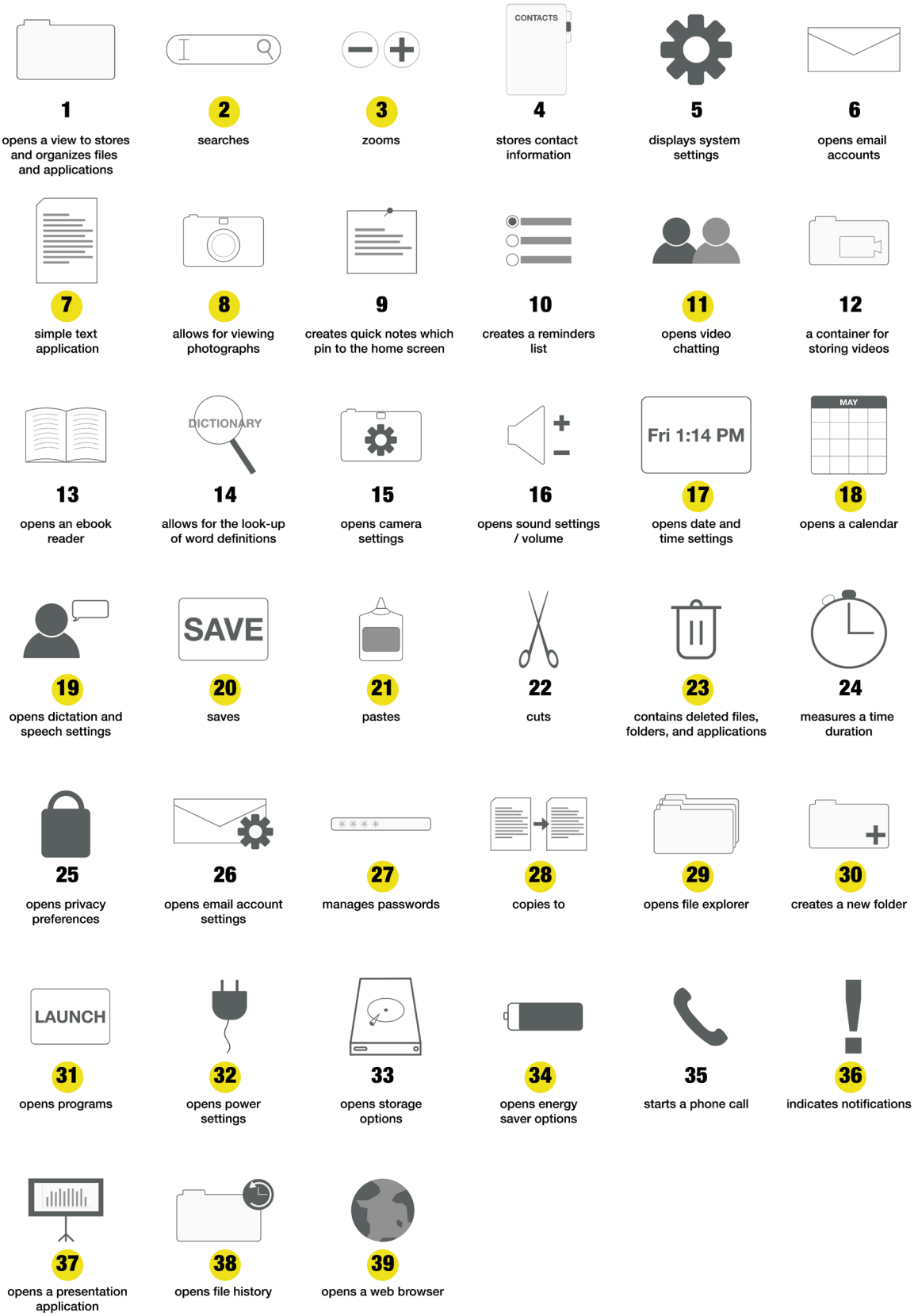

Fig. 6. The set of user-generated icon concepts emerging from our end-user elicitation study. Icons marked in yellow are new concepts different from the plausibly anachronistic icons we assembled (22 of 39 here are new).

Table 5
The referents each user-generated icon would trigger in a computing system. Columns describe whether the icon is of a new concept or of an anachronistic object, the median match score and the 1st and 3rd quartile on 1–7 Likert scale (higher is a better perceived symbol-referent match), and the median novelty score and the 1st and 3rd quartile on 1–3 Likert scale (1. I know where I have seen this icon. 2. I think I am familiar with this icon. 3. I have never seen this icon.).

Draw an icon that...	New Concept?	Median Match Score (1^{st}, 3^{rd})	Median Novelty Score (1^{st}, 3^{rd})
1. ...opens a view to store and organize files and applications		7 (5, 7)	3 (2, 3)
2. ...searches	✓	5.5 (5, 6)	2 (2, 3)
3. ...zooms	✓	7 (6, 7)	3 (3, 3)
4. ...stores contact information		6 (5, 7)	2 (2, 3)
5. ...displays system settings		6 (5.5, 7)	3 (3, 3)
6. ...opens email accounts		6 (5, 7)	1.5 (1, 2)
7. ...is a simple text application	✓	7 (6.5, 7)	3 (3, 3)
8. ...allows for the viewing of photographs	✓	5 (5, 6)	3 (2, 3)
9. ...creates quick notes which pin to the home screen		5.5 (4.25, 6)	1 (1, 2.75)
10. ...creates a reminders list		6 (5, 6.75)	3 (2, 3)
11. ...opens video chatting	✓	4 (4, 5)	1 (1, 1)
12. ...is a container for storing videos		6 (5, 6.75)	2 (1.25, 2)
13. ...opens an ebook reader		7 (6, 7)	2 (2, 3)
14. ...allows for the look-up of word definitions		5.5 (4.25, 6)	2 (1.25, 2.75)
15. ...opens camera settings		6 (5, 7)	2 (1, 3)
16. ...opens sound settings / volume		7 (6, 7)	3 (3, 3)
17. ...opens date & time settings	✓	6 (5, 6)	1 (1, 2.5)
18. ...opens a calendar	✓	7 (6, 7)	3 (2, 3)
19. ...opens dictation and speech settings	✓	5 (4, 6)	1.5 (1, 2)
20. ...saves	✓	5 (4, 7)	2 (1, 2.75)
21. ...pastes	✓	5 (3.5, 6)	1 (1, 2)
22. ...cuts		7 (6, 7)	3 (2, 3)
23. ...contains deleted files, folders, and applications	✓	7 (5, 7)	3 (2, 3)
24. ...measures a time duration		5 (5, 6)	3 (3, 3)
25. ...opens privacy preferences		7 (6, 7)	3 (2, 3)
26. ...opens email account settings		7 (5, 7)	1.5 (1, 2)
27. ...manages passwords	✓	6 (4, 6)	1 (1, 2)
28. ...copies to	✓	6 (4, 6)	1 (1, 2)
29. ...opens file explorer	✓	5.5 (4.25, 6)	2 (1, 2)
30. ...creates a new folder	✓	6 (5, 7)	2 (2, 3)
31. ...opens programs	✓	5.5 (4.75, 6.25)	1 (1, 1.5)
32. ...opens power settings	✓	6 (6, 7)	2 (1, 2.75)
33. ...opens storage options		6 (6, 7)	2 (1, 3)
34. ...opens energy saver options	✓	5 (5, 6)	2 (2, 3)
35. ...starts a phone call		6 (5, 7)	2 (2, 3)
36. ...indicates notifications	✓	6 (6, 7)	3 (2, 3)
37. ...opens a presentation application	✓	6 (5, 7)	2 (1, 3)
38. ...opens file history	✓	5 (3, 5)	1 (1, 1.5)
39. ...opens a web browser	✓	6 (5, 7)	3 (2, 3)

referents they trigger, their plausibly anachronistic icons, and the user-generated icon elicited from our 30 young-adult participants.

5.2.2. Referent accuracy

With 30 young-adult participants and the 39 plausibly anachronistic icons in Fig. 2, we obtained $30 \times 39 = 1{,}170$ attempted icon identifications for the computer functions those icons trigger. After grouping referents by similarity, nine referents of 39 (23.1%) were *not* identified correctly from their anachronistic icons, *i.e.,* the referent proposed with the highest consensus by our participants did not match the referent currently triggered by the icons in the actual computing systems from which those icons were extracted. The referents that were not identified correctly were: #7. "simple text application;" #15. "open camera settings;" #19. "open dictation and speech settings;" #22. "paste;" #27. "open email account settings;" #31. "create a new folder;" #32. "open program;" #35. "open energy saver options;" and #37. "indicate notifications." In Table 8, we list the incorrectly identified icons, the referent with the highest consensus proposed by our participants, the number of participants who proposed that referent, the actual referent associated with the icon, and the number of participants who proposed the actual referent.

5.2.3. Experience of real-world objects

We also asked our 30 young-adult participants additional questions beyond having them identify computing functions from icon representations. We asked participants to identify the real-world objects depicted by the plausibly anachronistic icons in Fig. 2, what those objects are used for, and when they last interacted with them. Twenty-five objects out of 38 (65.8%; icons #2. "search" and #3. "zoom" were the same object, a magnifying glass) were identified correctly by all 30 participants, and the remaining 13 objects (34.2%) were identified correctly by an average of 86% (SD=11%) of participants.

We also asked participants if they had used the real-world object depicted by each icon. Twelve real-world objects in 16 icons (four icons shared the same object, a folder) were, in fact, used by all of our young-adult participants. (And by extension, 22 objects had never been used by some participants.) The real-world objects that had been used were: folders, pens and paper, print photographs, paper sticky notes, paper books, scissors, a recycling bin, a padlock, a paper envelope, jagged metal keys, a compact disk, a wall plug, and a filament light bulb. It is fair to say that these real-world objects had, therefore, *not* become anachronistic to our young-adult participants yet. (One can speculate as to how many years from now, if ever, these real-world objects *might* become unfamiliar to young adults.)

By contrast, the real-world objects depicted by 22 of the 38 plausibly anachronistic icons had *never* been used by some of our participants. Fig. 10 shows the percentage of participants who had never used each of the real-world objects. On average, 25.4% (SD=19.6%) of all participants had never used these 22 objects. These objects were: a magnifying glass, an address book, a gear, a stamp, a reminder list, a video camera, a film strip, a paper dictionary, a dedicated camera, an audio speaker, a wall clock, a wall calendar, a microphone, a 3.5" floppy diskette, a physical clipboard, an analog stopwatch, a manila folder organizer, a computer hard drive, a land-line telephone, a bell, a podium, and a magnetic compass.

5.4. Identification of icons generated by young adults

To test the identifiability of the set of user-generated icons from our young adults (Fig. 6), we conducted an identification study using Ali *et al.'s* (Ali et al., 2019) end-user identification method with 60 online participants recruited from Amazon's Mechanical Turk (MTurk). We report on the referent agreement rates reached by the 60 participants and their accuracy in identifying the referent associated with each user-generated icon.

5.4.1. Referent agreement

We calculated the referent agreement for the referents collected in our online identification study using Eq. 1. There were 22 new concept icons and 17 icons that matched our plausibly anachronistic icons (Fig. 11). The mean referent agreement score for the new concept icons was 0.53 (SD=0.24); the mean referent agreement score for the plausibly anachronistic icons was almost identical at 0.54 (SD=0.24). The fact these two agreement scores were nearly equal suggests that perhaps

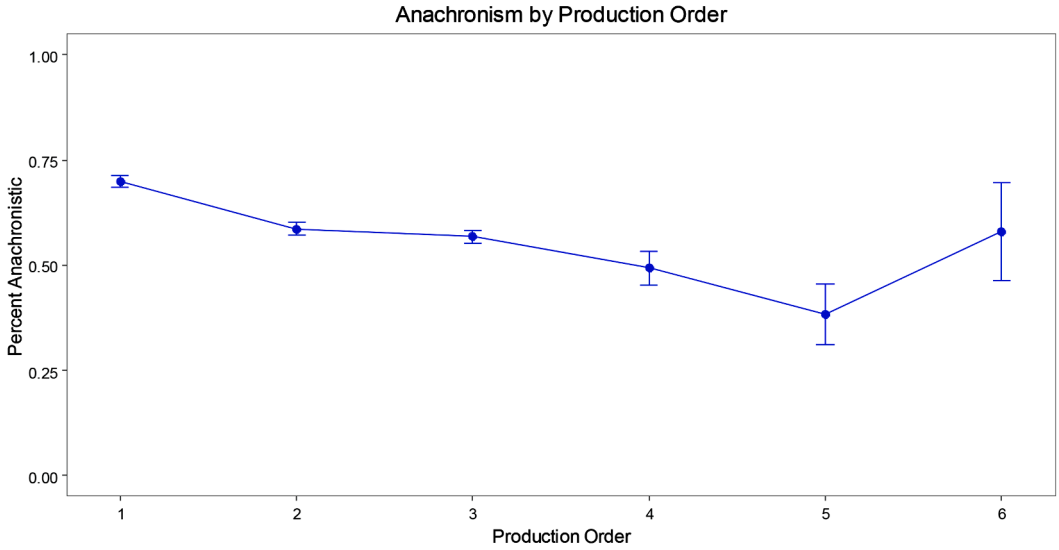

Fig. 7. The percent of icons that were anachronistic by production order Higher percentage is more likely to be anachronistic, while lower is less likely Error bars represent ±1 standard error.

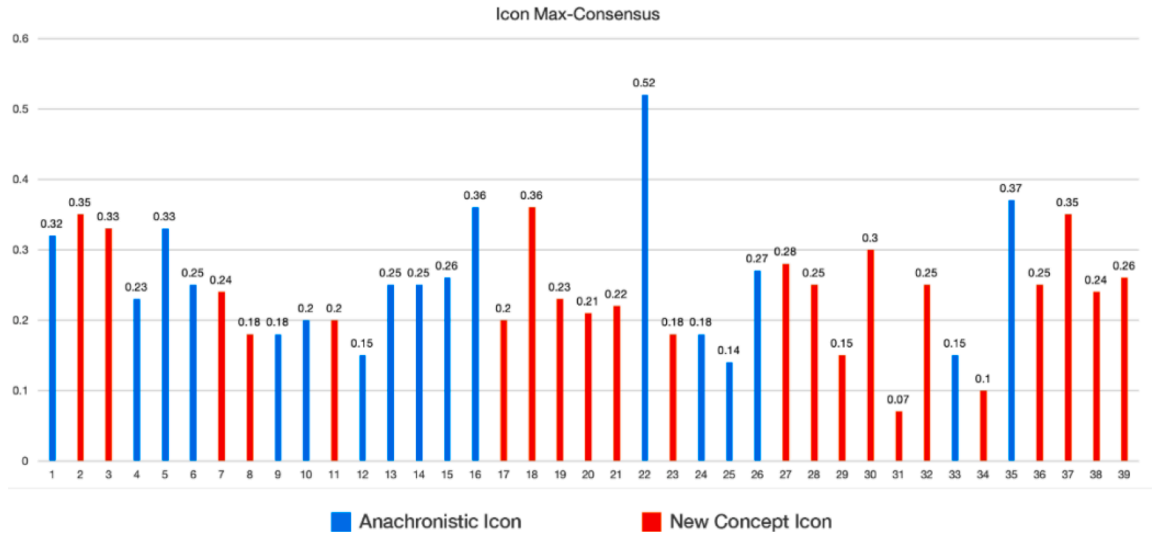

Fig. 8. Max-consensus scores for the 39 icons elicited from the 30 young-adult participants. The scores range from [0–1], with 1.00 being total agreement, i.e., all participants proposed the same icon. The blue bars represent anachronistic icons, and the orange bars represent new concept icons (22 of 39 are new). Refer to Figure 6 for the 39 user-generated icons.

the user-generated icons representing new concepts are no more, and no less, identifiable than their plausibly anachronistic predecessors.

5.4.2. Referent accuracy

We collected a total of 2,340 referent proposals for our 39 user-generated icons from 60 new online participants. A total of 61.6% (1,441) of the collected referent proposals were correct—*i.e.,* the participants were able to identify the actual referent the icon should trigger. For each user-generated icon, we found the referent with the highest consensus among our 60 new participants. The participants from MTurk were able to correctly identify 34 of the 39 user-generated icons (87.2%). Five user-generated icons had an incorrect referent with the highest consensus (12.8%). The five icons identified incorrectly were: #11. "open video chat;" #19. "open dictation and speech settings;" #24. "measure time duration;" #25. "open privacy preferences;" and #39. "open web browser." Table 9 lists the icons, the referent with the highest consensus resulting from our identification study, the correct referent associated with the icon, the number of participants who proposed the referent with the highest consensus, and the number of participants who proposed the correct referent.

6. A final set of icons

To recap, these were the parts of our multi-part study: (1) a lab-based elicitation study with 30 young-adult participants eliciting icon sketches for 39 computing functions; (2) a lab-based identification study of 38 plausibly anachronistic icons from current operating systems and applications; and (3) a crowdsourced identification study of 39 user-generated icons with 60 new online participants.

Drawing on the results of these studies, we have crafted arguably an improved set of icons for 39 computing functions (Fig. 12). By "improved," we mean icons that have high identifiability of their functions regardless of whether they represent a plausibly anachronistic object or not. The 39 computing functions are currently portrayed by at

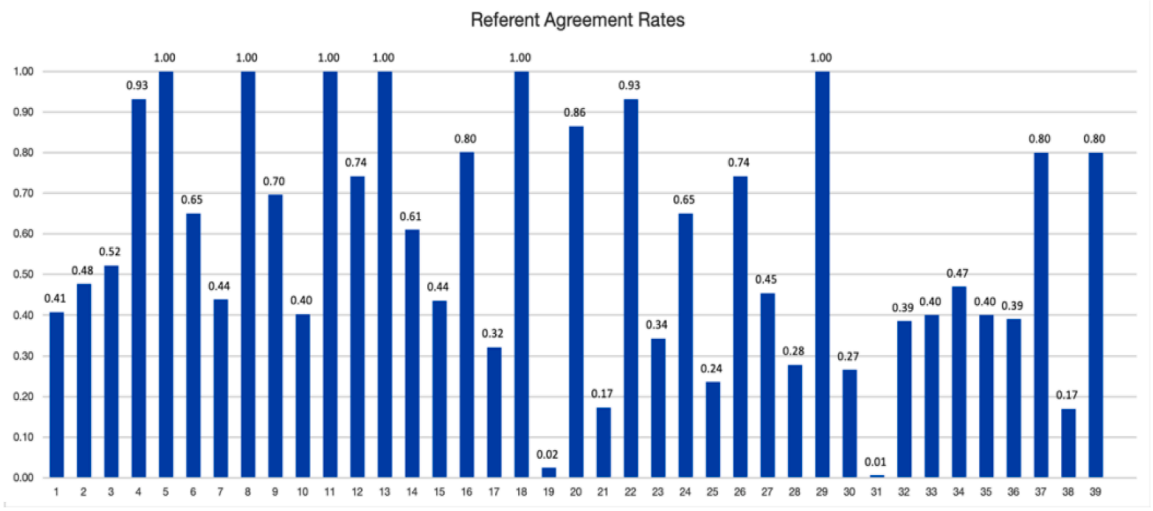

Fig. 9. The referent agreement scores for each of the 39 plausibly anachronistic icons from Figure 2 in our in-lab identification study with our 30 young-adult participants. Eight of the icons had agreement at 0.90 or above.

Table 6
Eight icons with agreement scores above 0.90, the referents they trigger, and the participant-generated icon created for these referents.

No.	Referent: "An icon that…"	Anachronistic Icon	Elicited Icon
4	Stores contact information		
5	Displays system settings		
8	Allows for the viewing of photographs		
11	Opens video chatting		
13	Opens an ebook reader		
17	Opens date & time settings		
22	Cuts		
29	Opens file explorer		

least some icons that depict anachronistic objects. Twenty-seven of the real-world objects depicted by these icons were *never* used by a quarter of our young-adult participants.

This final set of 39 user-generated icons has three categories: (1) *Anachronistic by Elicitation* icons are icons generated in our elicitation study that represent anachronistic objects; (2) *New Concept* icons are of new concepts that resulted from our end-user elicitation study; and (3) *Anachronistic by Identification* icons are those that had new icon concepts as a result of the elicitation study, but they were not identified correctly in our online identification study—hence, we chose the anachronistic icons rather than the new concepts.

The *Anachronistic by Elicitation* subset has 15 icons for the following referents: #1. "opens a view to store and organize files and applications;" #4. "store contact information;" #5. "display system settings;" #6. "open email account settings;" #9. "create quick notes that pin to the home screen;" #10. "create a reminder list;" #12. "open a container for storing videos;" #13. "open an e-book;" #14. "allow for the lookup of word definitions;" #15. "open camera settings;" #16. "open sound settings;" #22. "cut;" #26. "open email account settings;" #33. "open storage options;" and #35. "start a phone call."

The *Anachronistic by Identification* list has four icons for the following referents: #11. "open video chat;" #24. "measure a time duration;" #25. "open privacy preferences;" and #39. "open a web browser."

Finally, the *New Concept* list contains icons for the remaining 20 computing functions, all of which were identified correctly by our 60 new online participants in the identification study. One notable exception is the icon for referent #19—"open dictation and speech settings"—that had a new concept icon as a result of the elicitation study. However, the original anachronistic icon for referent #19 was not identified correctly in the lab-based identification study, and the new

Table 7
Seven icons with agreement scores lower than 0.30, the referents they trigger, and the user-generated icon created for these referents.

No.	Referent: "An icon that…"	Anachronistic Icon	Elicited Icon
19	Opens dictation and speech settings		
21	Paste		
25	Opens privacy preferences		
28	Copies to		
30	Creates a new folder		
31	Opens program		
38	Opens file history		

Table 8
Nine of 39 plausibly anachronistic icons were identified incorrectly by our 30 young-adult participants. The table lists the identified referent, the number of participants who proposed it, the actual referent, and the number of participants who proposed that.

No.	Icon	Identified Referent	No. Participants (N=30)	Actual Referent	No. Participants (N=30)
7		Take notes	17	is a simple text application	7
15		open webcam	17	opens camera settings	6
19		record voice	14	opens dictation and speech settings	1
22		copy	9	pastes	5
27		send email	24	opens email account settings	2
31		open folder	11	creates a new folder	8
32		install software	9	opens programs	1
35		control brightness	19	opens energy saver options	1
37		set an alarm	14	indicates notifications	10

concept icon was not identified correctly in the online identification study, either. The reason we include the new concept icon in our final icon set, rather than its anachronistic counterpart, is that the new icon had a higher referent agreement rate of 0.57 compared to the 0.02 referent agreement rate the anachronistic icon received. Despite this new concept icon being unsuitable for the referent, it is a more identifiable icon on its own. Finding a more reliably identifiable icon for referent #19 remains a possibility for a future end-user elicitation study.

We established a set of codes when we conducted our open-coding analysis on the entire set of 3,590 icon sketches that we elicited from our 30 young-adult participants in our lab-based elicitation study. We used these codes on our final set of icons (Table 10). Most icons in our final set (77%) represent a physical object. Forty-one percent of the icons utilize the desktop metaphor. About a quarter (26%) of the icons incorporating GUI elements had a "Meta GUI" code. Finally, about half (49%) of the icons were of plausibly anachronistic objects from our

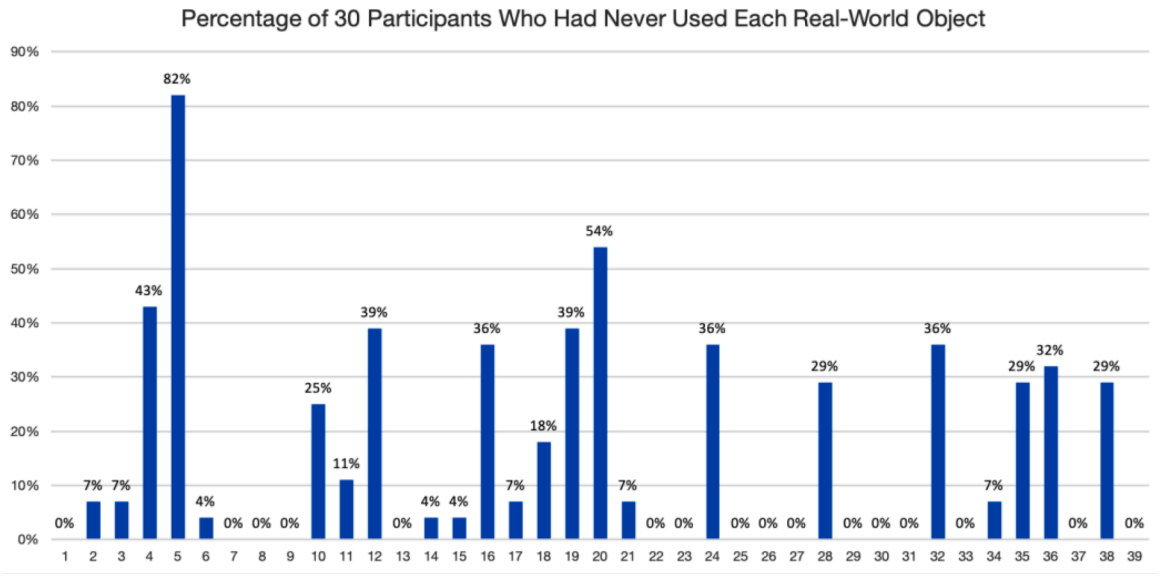

Fig. 10. The percentage of 30 participants who had never used the real-world object shown in each one of the 39 plausibly anachronistic icons. For example, 82% of participants had never used icon #5, which is a mechanical gear cog. See Figure 2 for the set of plausibly anachronistic icons.

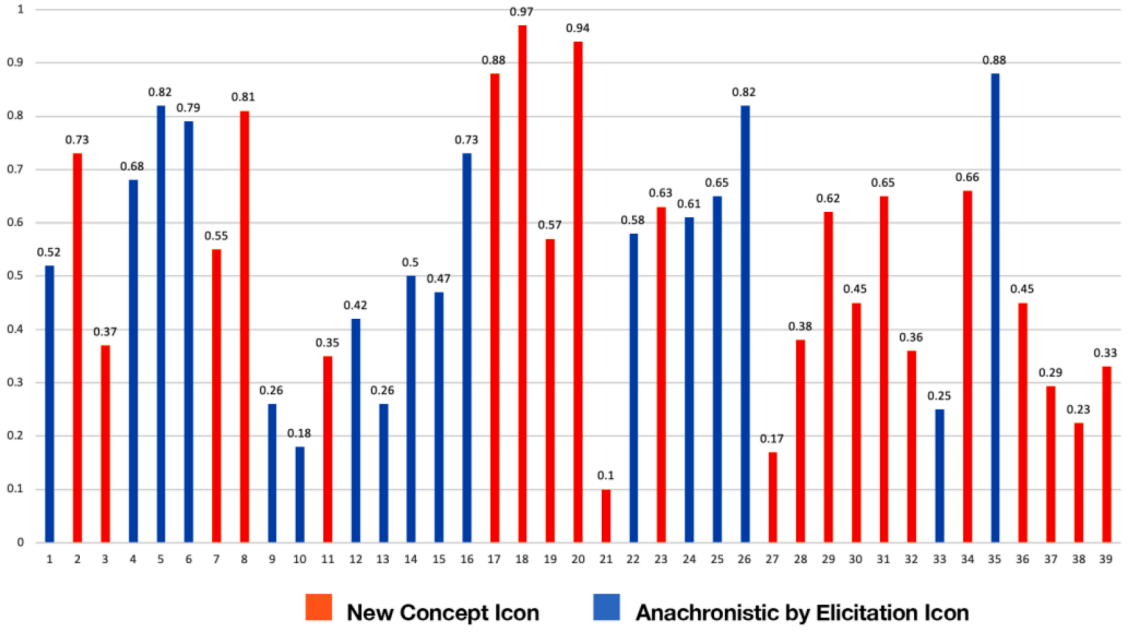

Fig. 11. The referent agreement scores for each of the 39 user-generated icons in Figure 6 in our online identification study with our 60 new participants. Blue bars represent the 17 icons that remained plausibly anachronistic; orange bars represent the 22 new concept icons.

original set (Fig. 2).

7. Discussion

To discuss our findings in context, we revisit the three research questions we outlined in the introduction of this article. We looked over current desktop operating systems and applications and assembled a list of 39 icons that represented plausibly anachronistic objects. Based on those icons, we set out to answer the following research questions: *What icons would young adults sketch to trigger each one of the 39 computer functions? How familiar are young adults with the objects represented in the 38 selected icons? How identifiable is a set of icons elicited from young adults?* We address each of these questions in turn. In addition, we highlight 20 anachronistic icons deserving of a redesign, we discuss the limitations of our work, and we identify promising directions for future work.

7.1. User-generated icons

We conducted open-coding analysis (Strauss and Corbin, 1997) on the entire set of 3,590 icon sketches we gathered in our lab-based elicitation study with 30 young-adult participants. The icon sketches proposed by participants were heavily influenced by physical objects. We found that 73% of all sketches represented a physical object. A similar number was found in the final set of 39 icons (Fig. 12), as 77% (30 icons) of these icons represented physical objects, too.

The use of alphanumeric characters, or "Text" as we categorized it in our coding scheme (Table 10), was present in 10 of 39 (26%) icons in our

Table 9
The five icons were identified incorrectly in our online identification study. An asterisk (*) indicates a new concept. The table lists the identified referent, the number of participants out of 60 who proposed the identified referent, the actual referent, and the number of participants who proposed that.

No.	Elicited Icon	Identified Referent	No. Participants (N=60)	Actual Referent	No. Participants (N=60)
11*		Open contacts	28	Open video chatting	21
19*		Open messages	45	Open dictation and speech settings	0
24		Show time	47	Measure a time duration	7
25		Lock computer	48	Open privacy preferences	0
39*		Open map	26	Open a web browser	22

final set. In four cases, the icons were entirely made up of text. These four icons were for the following referents: #17. "open date and time settings;" #20. "save;" #31. "open programs;" and #36. "indicate notifications." All four icons were identified correctly in our online identification study and had high referent agreement rates—above 0.60—except for referent #36, which still had a decent agreement rate of 0.45. This is not surprising, as text icons spell out the function they trigger. A blatant example of using text was for referent #20. "Save," which was only the outlined word "Save," and had a referent agreement rate—or identifiability score—of 0.94. Such a choice for "Save" stands in stark contrast to the 3.5" floppy diskette icon so prevalent in today's desktop interfaces, and begs the question of whether it is time for an overhaul of modern computer iconography.

For icons that were *Anachronistic by Elicitation*, none of them had a max-consensus score of 0.10 or less. In fact, five *Anachronistic by Elicitation* icons had max-consensus above 0.30, which were #1. "opens a view to store and organize files, and applications" (0.32), #5. "displays system settings" (0.33), #16. "opens sound settings/ volume" (0.36), #22. "cuts" (0.52), and #35. "starts a phone call" (0.37). Initially, we had considered the objects represented in these icons—folder, cog, speaker, scissors, landline handset—to be plausibly anachronistic. We dive into the anachronism of the objects depicted in our set of 38 plausibly anachronistic icons in the next subsection.

Five of the new concept icons had max-consensus scores above 0.30. They were #2. "searches" (0.35), #3. "zooms" (0.33), #18. "opens a calendar" (0.36), #30. "creates a new folder" (0.30), #37. "opens a presentation application" (0.35). However, there were two new concept icons with 0.10 or less max-consensus scores: #31. "opens programs" (0.07) and #34. "opens energy saver options" (0.10).

The elicited new concept icons seem to be a viable update for existing anachronistic icons in some cases, like the anachronistic icon #2. "searches," which was represented by a magnifying glass—two (7%) of our participants had never used a magnifying glass before. The icon had an identifiability score of 0.48 from our young adults, yet they elicited a new concept icon for it that had a 0.73 identifiability score. For icon #18. "opens a calendar," even though the anachronistic icon representing a desk calendar was identified correctly by all 30 participants and had a perfect referent agreement score of 1.00, it had a new concept icon that had similarly high referent agreement score of 0.97 in the online identification study. Given the fact that five (18%) of our young-adult participants had never used a desk calendar before, the updated digital calendar seems like a useful improvement for this icon. Icon #30. "creates a new folder" had the icon that represents a folder, an object that all our participants had used. The young adult participants updated this icon by incorporating the folder and adding a (+) sign to it. This update seems to have improved the icon given that its referent agreement score in the online identification study went up to 0.45.

New concept icons improved identifiability scores even when the consensus on what the new icon should be was low. This was the case for icons #31. "opens programs" and #34. "opens energy saver options"—both had low agreement scores in the elicitation study of 0.07 and 0.10 respectively. The new concept icons had higher identifiability than their anachronistic counterparts as #31 had an identifiability score of 0.01—even though the objects, a box and a compact disk, had been used by all of our participants. Identifiability using the new icon for referent #31 increased from 0.01 to 0.65. Icon #34 when represented by a light bulb had a 0.47 identifiability score that increased to 0.66 for the new icon using a battery.

On the other hand, there were cases where the young-adult participants elicited a new concept icon that was identified correctly in our online identification study, thus warranting its inclusion in our final set of icons (Fig. 12), but whose anachronistic counterpart was nonetheless more identifiable. An example is icon #37. "opens a presentation," which had an elicitation max-consensus score of 0.35, depicting a screen showing a bar graph. In our online identification study, this bar graph icon had a 0.29 identifiability score. The anachronistic icon for #37 was a podium, an object that had been used by all of our participants, and had high identifiability of 0.80. Another example is icon #3. "zooms," represented by a magnifying glass. Its new concept icon had two circles with the signs (-) and (+). This new concept icon had an identifiability score of 0.37. Its anachronistic counterpart, the magnifying glass, had an identifiability score among our young-adult participants of 0.52.

7.2. Anachronism

The results of our investigations culminated in a final set of icons informed by three user studies (Fig. 12). About half of the final set of icons—19 of 39—are what we initially thought of as a plausibly anachronistic icon. The number of anachronistic icons that ended up in the final set is consistent with the total number of anachronistic icons elicited in our lab-based elicitation study with our 30 young-adult participants, 1,965 of the 3,590 total icons (55%). Despite that our 30 young-adult participants identified all of the 38 real-world objects represented in the plausibly anachronistic icons—25 objects were identified correctly by all participants, and the remaining 13 were identified by more than half of the participants—all of our young-adult participants had only ever used 16 of these objects. On average, a quarter of our young adults had never used the objects depicted in 23 of the 39 anachronistic icons.

We further explore anachronism as it relates to our set of 39 icons and divide it into three categories: (1) *truly anachronistic icons* are those that represent objects that some of our young adults have not used and proposed new concept icons to trigger their functions; (2) *usable anachronistic icons* are those that represent objects that some of our young adults have not used but proposed icons that still represent the same anachronistic icon; (3) *not anachronistic icons* are those that we

Fig. 12. The final set of icons based on our elicitation and identification studies.

initially classified as anachronistic, but found that at least one of our young-adult participants had used the object depicted in the icon.

Truly anachronistic icons are those that represent objects that some of our young adults have not used and proposed new concept icons to trigger their functions. There were 13 truly anachronistic icons: #2. "searches," #3. "zooms," #11. "opens video chat," #17. "opens date and time settings," #18. "opens a calendar," #19. "opens dictation and speech settings," #20. "saves," #21. "pastes," #28. "copies to," #32.

Table 10
Analysis of our final set of icons.

Code	Icon Referent No.	N = 39
Object	1, 2, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 28, 29, 30, 32, 33, 34, 35, 37, 38, 39	30 (77%)
Metaphor		
Desktop	1, 4, 6, 7, 9, 13, 14, 21, 22, 24, 26, 28, 29, 30, 35, 38	16 (41%)
New Technology		0
Mobile		0
Anachronistic Practice		0
Computer Hardware	33	1 (3%)
Computer Action	10, 27, 28, 34	4 (10%)
Rare		0
Body Part	19	1 (3%)
Physical Activity		39 (1%)
Text	3, 4, 14, 16, 17, 18, 20, 30, 31, 36	10 (26%)
Shape		
Abstract		0
Design	7, 9, 10, 13, 28	5
Convention		(13%)
Nature		0
Brand		
Brand Name		0
Logo		0
Brand Influence		0
Meta GUI	2, 10, 18, 19, 27, 37	6 (15%)
Anachronistic object		
Existing icon		19 (49%)
New Anachronism		0

"opens power settings," #34. "opens energy saver options," #36. "indicates notifications," and #38. "opens file history." These new concept icons on average had a max-consensus rate of 0.25 in our elicitation study. The anachronistic icons had an average of 0.47 referent agreement in our in-lab identification study. Their *New Concept Icons* had an increased average referent agreement score of 0.54.

Usable anachronistic icons are those that represent objects that some of our young-adult participants had not used, but for which they proposed icons that still represented the same anachronistic icon. There were 10 *usable anachronistic icons* that depicted anachronistic objects which were: #4. "stores contact information," #5. "displays system settings," #6. "opens e-mail accounts," #10. "creates a reminders list," #12. "is a container for storing videos," #14. "allows for the lookup of word definitions," #15. "opens camera settings," #16. "opens sound settings / volume," #24. "measures a time duration," and #35. "starts a phone call." These *usable anachronistic icons* had an average max-consensus rate of 0.26. The in-lab and online identification referent agreement rates for these *usable anachronistic icons* were close at 0.66 and 0.61, respectively.

We called these icons *usable anachronistic icons* because even though young-adult participants had never used these objects, the anachronistic icons that depict these objects are still highly identifiable. For example, even though 82% of our participants had never used a cog in real life for icon #5. "displays system settings," all 30 participants were able to identify the correct function associated with the icon. Also, because icon #5 had a 0.82 referent agreement in the identification study with our 60 online participants, it seems that having a cog to "display system settings" is still a good icon choice. The same goes for icon #16. "open sound settings / volume," which is represented by a speaker. Even though 36% of our young adults never used such a speaker in real life, the referent agreement in the identification study with our young adults was 0.80; in our online identification study it was 0.73. For icon #35, which used a landline handset to represent the "starts a phone call" function, 29% of our participants had never used a landline in real life. The referent agreement rate in the identification study with our young adults was 0.40, less than half of that for our online identification study, which was 0.88.

For these three icons, we believe their high identifiability and the reason they remained *anachronistic by elicitation* could be attributed to the fact that these same icons represent the same functions in smartphone operating systems. So even though, on average, more than a quarter of our young adults had never used these three objects in real life, they still perceived them to be the best iconic representation of their functions.

Not anachronistic icons are those that we initially classified as anachronistic but found that at least one of our young-adult participants had used the object depicted in the icon. These 16 icons were: #1. "opens a view to store and organize files and applications," #7. "is a simple text application," #8. "allows for the viewing of photographs," #9. "creates quick notes which pin to the home screen," #13. "opens an ebook reader," #22. "cuts," #23, "contains deleted files, folders, and applications," #25. "opens privacy preferences," #26. "opens email account settings," #27. "manages passwords," #29. "opens file explorer," #30. "creates a new folder," #31. "opens programs," #33. "opens storage options," #37. "opens a presentation application," and #39. "opens a Web browser." For the *not anachronistic icons*, the average max-consensus score in the elicitation study was 0.24. The in-lab identification study had a referent agreement rate of 0.60, and the online referent agreement rate was 0.49.

We discovered that folders and scissors are items still being used by today's young adults: all of our 30 participants reported that they have used these physical objects. These two objects also work well as icons associated with the computing functions they represent: folders for #1. "opens a view to store and organize documents, files, and applications," and scissors for #22. "cuts." From our lab identification study, icon #1 had 0.41 referent agreement and icon #22 had 0.93 referent agreement. Their online identifiability scores were 0.52 and 0.58, respectively. These two icons are examples of icons that do not need to change due to a lack of identifiability.

Even though they were *not anachronistic*, 8 of 16 *not anachronistic icons* had *New Concept Icons*, which indicates, to us, that our computer iconography needs updating even if some of these icons represent physical objects we still use. These icons were: #7. "is a simple text application," #8. "allows for the viewing photographs," #23. "contains deleted files, folders, and applications," #27. "manages passwords," #29. "opens file explorer," #30. "creates a new folder," #31. "opens programs," and #37. "opens a presentation application."

7.3. Identifiability of the user-generated set of icons

The set of user-generated icons that resulted from our elicitation study (Fig. 6) was highly identifiable, at an average referent agreement rate—or identifiability score—of 0.54. In fact, despite the fact that in our online identification study we showed icons alone without any context or labels, only five icons' computing functions were *not* identified correctly. We reverted these icons to their original plausibly anachronistic icons in our final set of icons (Fig. 12) because their anachronistic counterparts were identified correctly in our lab-based identification study. A single exception was the icon for referent #19. "open speech and dictation settings." It was not identified correctly in either identification study. Perhaps this is because the function of opening speech and dictation settings is not as widely used as other functions.

The icons with the highest identifiability scores were #18. "opens a calendar" and #20. "save," with over 0.90 referent agreement rates. Icons #18 and #20 were *New Concept Icons* that included alphanumeric text, which explains the high identifiability scores. Other user-generated

icons with relatively high referent agreement scores of 0.80 and above included #17. “opens date and time settings,” which was a *New Concept Icon* from the *truly anachronistic* group that had a 0.88 identifiability score. Icon #35. “starts a phone call,” a *usable anachronistic icon*, had a 0.88 identifiability score as well. These icons with high identifiability scores support the motivation for this work and demonstrate that user-generated icons are viable updates to anachronistic icons.

The icon with the lowest identifiability score (0.10) was #21. “pastes,” represented by a glue bottle, which was a *New Concept Icon* with 0.22 symbol agreement rate among the young-adult participants. The glue bottle replaced the clipboard icon, which had a 0.17 identifiability score. Clearly, further design work is needed to create a highly identifiable icon for the "paste" function.

Other icons with relatively low identifiability scores were #27. “manages passwords,” at 0.17, and #10. “creates a reminders list,” at 0.18. Icon #27 was a *New Concept Icon*—a password field that could be found on websites—elicited with 0.28 max-consensus. It replaced a set of metal keys, not unfamiliar anachronistic objects, which had a 0.47 identifiability score. This is a case where the original icon (the keys)—which we had initially classified as anachronistic—was more identifiable than the user-generated *New Concept Icon*, and probably should remain unchanged. As for icon #10, it had a *usable anachronistic icon*, that of a bulleted list. Our young adults were able to identify it with a 0.40 referent agreement rate. The young-adult participants had a max-consensus of 0.22 when eliciting this anachronistic icon, and 25% of them had never used a physical “to do” checklist. In this case, we might attribute the lowered identifiability score to our own simplistic illustration of the icon. (Compare icon #10 in Fig. 2 and Fig. 5.)

7.4. Icons due for a redesign

Well designed icons are necessary for usability (Chen et al., 2013, McKnight and Read, 2009). In this section, we highlight 20 icons currently found in desktop operating systems and applications that could benefit from an update as a result of this work. First, *truly anachronistic icons* need an update immediately as they represent anachronistic objects that young adults have never used and for which they proposed new concept icons. These *truly anachronistic icons* are: #2. “searches,” #3. “zooms,” #11. “opens video chat,” #17. “opens date and time settings,” #18. “opens a calendar,” #19. “opens dictation and speech settings,” #20. “saves,” #21. “pastes,” #28. “copies to,” #32. “opens power settings,” #34. “opens energy saver options,” #36. “indicates notifications,” and #38. “opens file history.” These icons were effectively redesigned in our elicitation study, the results appearing in Fig. 5 and Fig. 12.

The second set of icons we think could benefit from an update is the *usable anachronistic icons*. While they are currently identifiable icons for the functions they trigger, they still represent objects that young adults no longer use. As a result, they might be headed towards “truly anachronistic” status with the passage of more time. These icons are: #5. “displays system settings,” #4. “stores contact information,” #10. “creates a reminders list,” #12. “is a container for storing videos,” #16. “opens sound settings / volume,” #24. “measures a time duration,” and #35. “starts a phone call.” (Note that we highlighted icons #5, #16, #35 in Section 7.2 where we discussed *usable anachronistic icons* due to their high identifiability scores.)

Our final set of icons (Fig. 12) could serve as inspiration for redesigning these icons. We generated this final set of icons to exhibit the results of our multi-part study, but not as a definitive redesign of the 39 plausibly anachronistic icons that served as the basis for this work. Future elicitation studies with larger more diverse populations would be beneficial in informing the design of the 20 icons we highlighted in this section. Perhaps a distributed elicitation study—like in the work of Ali et al. (2019)— involving end users of different backgrounds could inform the design of these icons to be identifiable to a wide population of potential users. In addition, platform-specific considerations and guidelines would be important to consider in any redesign effort.

7.5. Limitations and future work

As with any study, and especially in the case of three studies, there are certain limitations. Despite the diversity of our 30 young-adult participants’ countries of origin, our participants were all people attending a major university in the United States. Most of our participants used technology quite heavily: 6–10 hours per day using a computer, 3–5 hours per day using a mobile device, and less than 2 hours per day using certain physical objects. It would be interesting to replicate this study in the future with a larger, more diverse set of young users. Additionally, an elicitation study with 30 college students is not a large enough sample on which to base an entire redesign of, say, Microsoft’s or Apple’s operating systems icons. However, this work provides initial evidence that suggests preliminary directions for such design work.

Future work could reverse our approach by having young adults identify icons elicited from older adults, or explore the differences in the guessability of icons created by culturally diverse populations.

We utilized one of three possible techniques from Morris et al. (2014) to reduce legacy bias in our elicitation study, the “production principle,” which asked our participants to sketch as many icons as they could imagine, rather than only one. We also discovered some statistical support that employing this principle did, indeed, reduce the chances for anachronism (Fig. 7). A future study could utilize Morris *et al.*’s other legacy bias reduction principles such as “priming” or “partners.” For example, the plausibly anachronistic icons in Fig. 2 could be shown as *prohibited* examples that participants would not be allowed to propose. Or, participants could work together to brainstorm additional icon possibilities. Our results showed that 55% of the 3,590 icons we collected and 19 of the 39 icons we had in our final icon set were still plausibly anachronistic. By using additional techniques to prioritize novelty, we might reduce potential anachronism further. The viability of an icon set resulting from such a study would be a question best answered by an identification study, which could compare that set of icons to the final set of icons we derived in this work (Fig. 12).

As noted above, none of our computing functions or icons were perceived in context by participants. Rather, they were always described or depicted in isolation, apart from any graphical user interface in which they might be situated. Further validation of any icon set requires contextualizing that set in real user interfaces, whether visual mockups, functional prototypes, or even commercial products. Contextualized guessability studies could ask participants to predict the behavior of computing systems before icons are selected, which would provide insight into contextualized identifiability. Similarly, if participants are ever surprised at the computing function that results from using an icon, they can suggest a different, perhaps better, alternative.

8. Conclusion

In this work, we sought to understand young-adult technology users’ perceptions of yesterday’s and today’s computer iconography. To this end, we conducted a multi-part study that resulted in a set of 39 user-generated icons for common computing functions. Half of these icons were new concepts and the other half remained anachronistic, indicative of objects or concepts from an earlier era but which nonetheless still have meaning for today’s young-adult computer users. Our work provides insight into young adults’ high reliance on physical objects to depict computer iconography. We provide a taxonomy of computer iconography and highlight 20 anachronistic icons in need of redesign. It is our hope that researchers can utilize the findings from our work to better understand the mental models and perceptions of today’s young-adult computer users. We also hope that user interface designers can employ our findings to direct their efforts at updating computer icons for the next generation of computer users. It is no longer the floppy disk that means “Save.” Rather, it is “Save” that is represented by a small abstract

square.

Credit author statment

Abdullah X. Ali: Conceptualization, methodology, draft preparation, writing, reviewing and editing, data collection, data analysis, graphics creation, team coordination.

Erin MCAWEENEY: Initial draft preparation, data collection, data analysis.

JACOB O. WOBBROCK: Project idea, methodology, statistics, data analysis, writing, reviewing and editing, project support.

Declaration of Competing Interest

The authors of article have no conflict of interest to declare.

Acknowledgments

The authors thank Huiru Luo for help on this project. This research was supported in part by Microsoft, the University of Washington Information School, and the National Science Foundation under grant IIS-1702751. Any opinions, findings, conclusions or recommendations expressed in our work are those of the authors and do not necessarily reflect those of any supporter.

References

Ali, Abdullah X., Morris, Meredith Ringel, Wobbrock, Jacob O., 2018. Crowdsourcing Similarity Judgments for Agreement Analysis in End-User Elicitation Studies. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18* (UIST '18), pp. 177–188. https://doi.org/10.1145/3242587.3242621.

Ali, Abdullah X, Morris, Meredith Ringel, Wobbrock, Jacob O., 2019. Crowdlicit: A System for Conducting Distributed End-User Elicitation and Identification Studies. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3290605.3300485, 255.

Chen, Chien-Hsiung, Hsiao, Wen-Hsin, Chen, Shih-Chieh, Kang, Yen-Yu, 2013. Usability study of icon designs with social network functions. In: *Proceedings of the 15th international conference on Human Interface and the Management of Information: information and interaction for health, safety, mobility and complex environments - Volume Part II* (HCI International'13). https://doi.org/10.1007/978-3-642-39215-3_42, 355–362.

Collins, Belinda L., Lerner, Neil D., 1982. Assessment of Fire-Safety Symbols. Human Factors 24 (1), 75–84. https://doi.org/10.1177/001872088202400108.

Connell, Sabrina, Kuo, Pei-Yi, Liu, Liu, Piper, Anne Marie, 2013. A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In: Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13, pp. 277–280. https://doi.org/10.1145/2485760.2485823.

Findlater, Leah, Lee, Ben, Wobbrock, Jacob, 2012. Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12, pp. 2679–2682. https://doi.org/10.1145/2207676.2208660.

Freeman, Dustin, LaPierre, Nathan, Chevalier, Fanny, Reilly, Derek, 2013. Tweetris: a study of whole-body interaction during a public art event. In: Proceedings of the 9th ACM Conference on Creativity & Cognition - C&C '13, p. 224. https://doi.org/10.1145/2466627.2466650.

Gheran, Bogdan-Florin, Vanderdonckt, Jean, Vatavu, Radu-Daniel, 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In: Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18, pp. 623–635. https://doi.org/10.1145/3196709.3196741.

Good, Michael D., Whiteside, John A., Wixon, Dennis R., Jones, Sandra J., 1984. Building a user-derived interface. Communications of the ACM 27 (10), 1032–1043. https://doi.org/10.1145/358274.358284.

Ho, Chun-Heng, Hou, Kai-Chun, 2015. Exploring the Attractive Factors of App Icons. KSII Transactions on Internet and Information Systems 9 (6). https://doi.org/10.3837/tiis.2015.06.016.

Holm, Sture, 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 6 (2), 65–70.

Johnson, J., Roberts, T.L., Verplank, W., Smith, D.C., Irby, C.H., Beard, M., Mackey, K., 1989. The Xerox Star: a retrospective. Computer 22 (9), 11–26. https://doi.org/10.1109/2.35211.

Mauney, Dan, Howarth, Jonathan, Wirtanen, Andrew, Capra, Miranda, 2010. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10, p. 4015. https://doi.org/10.1145/1753846.1754095.

Mavri, Aekaterini, Loizides, Fernando, Zantides, Evripides, 2016. Communicating content: development and evaluation of icons for academic document triage through visualisation and perception. Behaviour & Information Technology 35 (9), 758–780. https://doi.org/10.1080/0144929X.2016.1194478.

McAweeney, Erin, Zhang, Haihua, Nebeling, Michael, 2018. User-Driven Design Principles for Gesture Representations. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, pp. 1–13. https://doi.org/10.1145/3173574.3174121.

McKnight, Lorna, Read, Janet C., 2009. Designing the "record" button: using children's understanding of icons to inform the design of a musical interface. In: *Proceedings of the 8th International Conference on Interaction Design and Children* (IDC '09), pp. 258–261. https://doi.org/10.1145/1551788.1551847.

Modanwal, Gourav, Sarawadekar, Kishor, 2018. A Gesture Elicitation Study with Visually Impaired Users. In: HCI International 2018 – Posters' Extended Abstracts. Springer International Publishing, Cham, pp. 54–61. https://doi.org/10.1007/978-3-319-92279-9_7. Constantine Stephanidis (ed.).

Morris, Meredith Ringel, 2012. Web on the wall: insights from a multimodal interaction elicitation study. In: Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS '12, p. 95. https://doi.org/10.1145/2396636.2396651.

Morris, Meredith Ringel, Danielescu, Andreea, Drucker, Steven, Fisher, Danyel, Lee, Bongshin, schraefel, c., Wobbrock, Jacob O., 2014. Reducing legacy bias in gesture elicitation studies. interactions 21 (3), 40–45. https://doi.org/10.1145/2591689.

Nebeling, Michael, Huber, Alexander, Ott, David, Norrie, Moira C., 2014. Web on the Wall Reloaded: Implementation, Replication and Refinement of User-Defined Interaction Sets. In: Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces - ITS '14, pp. 15–24. https://doi.org/10.1145/2669485.2669497.

Obaid, Mohammad, Häring, Markus, Kistler, Felix, Bühling, René, André, Elisabeth, 2012. User-Defined Body Gestures for Navigational Control of a Humanoid Robot. In: Social Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 367–377. https://doi.org/10.1007/978-3-642-34103-8_37. Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons and Mary-Anne Williams (eds.).

Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. User-Defined Gestures for Augmented Reality. IFIP Conference on Human-Computer Interaction. 2013. pp 282-299. https://link.springer.com/chapter/10.1007/978-3-642-40480-1_18.

Smith, David Canfield, 1977. Pygmalion: A Computer Program to Model and Stimulate Creative Thought. Retrieved November 5, 2019 from. https://link.springer.com/book/10.1007/978-3-0348-5744-4.

Smith, D.C., Irby, C., Kimball, R. Verplank, B., Harslem, E. (1982). Designing the Star user interface. BYTE Magazine, April 1982, pp. 242-282. https://tech-insider.org/star/research/acrobat/8204.pdf.

Stiratelli, Robert, Laird, Nan, Ware, James H., 1984. Random-Effects Models for Serial Observations with Binary Response. Biometrics 40 (4), 961–971. https://doi.org/10.2307/2531147.

Strauss, Anselm, Corbin, Juliet M., 1997. Grounded Theory in Practice. SAGE.

Tan, Yanke, Ho Yoon, Sang, Ramani, Karthik, 2017. BikeGesture: User Elicitation and Performance of Micro Hand Gesture As Input for Cycling. In: (CHI EA '17), pp. 2147–2154. https://doi.org/10.1145/3027063.3053075.

Tsandilas, Theophanis, 2018. Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation. ACM Transactions on Computer-Human Interaction 25 (3), 1–49. https://doi.org/10.1145/3182168.

Vatavu, Radu-Daniel, Wobbrock, Jacob O., 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, pp. 1325–1334. https://doi.org/10.1145/2702123.2702223.

Vatavu, Radu-Daniel, Wobbrock, Jacob O., 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, pp. 3390–3402. https://doi.org/10.1145/2858036.2858228.

Ware, Colin, 2010. Visual Thinking: For Design. Morgan Kaufmann. Retrieved November 13, 2019 from. https://www.researchgate.net/profile/Colin_Ware/publication/200027485_Visual_Thinking_For_Design/links/53ce927c0cf2aada06e6a4b4.pdf.

Wobbrock, Jacob O., Htet Aung, Htet, Rothrock, Brandon, Myers, Brad A., 2005. Maximizing the guessability of symbolic input. In: CHI '05 extended abstracts on Human factors in computing systems - CHI '05. https://doi.org/10.1145/1056808.1057043, 1869.

Wobbrock, Jacob O., Morris, Meredith Ringel, Wilson, Andrew D., 2009. User-defined gestures for surface computing. In: Proceedings of the 27th international conference on Human factors in computing systems - CHI 09. https://doi.org/10.1145/1518701.1518866, 1083.

Yan, Rushan, 2011. Icon Design Study in Computer Interface. Procedia Engineering 15, 3134–3138. https://doi.org/10.1016/j.proeng.2011.08.588.

Zhao, Nanxuan, Kim, Nam Wook, Herman, Laura Mariah, Pfister, Hanspeter, Lau, Rynson W.H., Echevarria, Jose, Bylinskii, Zoya, 2020. ICONATE: Automatic Compound Icon Generation and Ideation. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), pp. 1–13. https://doi.org/10.1145/3313831.3376618.

Gilmour, A.R., Anderson, R.D., Rae, A.L., December 1985. The analysis of binomial data by a generalized linear mixed model. Biometrika 72 (3), 593–599.