Regulating Access to Adult Content (with Privacy Preservation)

Karen Renaud & Joseph Maguire*
School of Computing Science, University of Glasgow, Glasgow, UK
{karen.renaud,joseph.maguire}@glasgow.ac.uk

ABSTRACT

In the physical world we have well-established mechanisms for keeping children out of adult-only areas. In the virtual world this is generally replaced by self declaration. Some service providers resort to using heavyweight identification mechanisms, judging adulthood as a side effect thereof. Collection of identification data arguably constitutes an unwarranted privacy invasion in this context, if carried out merely to perform adulthood estimation. This paper presents a mechanism that exploits the adult's more extensive exposure to public media, relying on the likelihood that they will be able to recall details if cued by a carefully chosen picture. We conducted an online study to gauge the viability of this scheme. With our prototype we were able to predict that the user was a child 99% of the time. Unfortunately the scheme also misclassified too many adults. We discuss our results and suggest directions for future research.

ACM Classification Keywords

H.1.2 User/Machine Systems: Human Factors

Author Keywords

Children; Protection; Access

INTRODUCTION

Automatically regulating access to adult content is challenging, especially if privacy is preserved [9]. The aim is to permit access to adults, but not to minors. In the physical world we effortlessly categorise people, in terms of age [24]. This makes it relatively easy for us to judge adulthood. When we are not sure, and being sure is a legal requirement, we may resort to identification. In the digital world, age-group classification without identification is challenging. While solutions exist (eg. German e-ID), they have not yet been widely accepted [19]. The concern is that society seems to be sleepwalking into using biometrics for this purpose [33]. Such widespread identification offsets the adult content provider's legal responsibility, but it also gives them a wealth of potentially sensitive information, something they might well abuse [49].

There is increasing concern from authorities across the world [33, 15], that children are accessing adult content online. Recent reports indicate that approximately 200,000 minors accessed age-inappropriate content in a single month [47], with the youngest being six years of age. Indeed, many minors claimed to have encountered pornographic images before the age of twelve [2]. The impact of such unrestricted consumption on developing minds should not be underestimated [11, 36, 47]. The existing mechanisms, are either ineffective or intrusive [51].

This paper explores a primary difference between adults and children that could be used to test for adulthood. Adults have lived longer and have accumulated more knowledge over the course of their lifetime [42, 38]. We propose to exploit these *knowledge differences* to achieve adulthood estimation without invading privacy. Our mechanism exploits this difference by testing the ability to label specially-chosen images correctly, essentially implementing a privacy-preserving access control mechanism. We first explain how we identified suitable images, and then how we used these images to judge the adulthood of experimental subjects in two iterations. We conclude by discussing our outcomes and presenting our suggestions for future developments. In summary, this paper's contributions are:

- The concept of knowledge-based regulation of access to adult content;
- Proven accuracy of the scheme based on two studies, using empirically validated images to perform adulthood estimation;
- Discussion of the implementation challenges and proposals for future work.

RELATED WORK

We first review existing mechanisms and then explain why we deployed images to test knowledge.

Existing Techniques

Parents use a variety of techniques to prevent their children from accessing adult content [33]: (1) education and

*School of Computing Science

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

supervision, (2) automated filtering on devices and/or (3) relying on content providers to prevent access.

Education alone is not entirely effective [30], judging by its efficacy in other contexts [22]. It could prove useful in conjunction with supervision. Nevertheless, while supervision may seem sensible it may have unintended consequences. Chirkov *et al.* argue that adolescents need to be allowed to express their autonomy, competence and relatedness in order to successfully come of age [10]. This suggests that parents and teachers should gradually relax control rather than tightening their grip if the child is to become a responsible adult.

Automated blocking and filtering of adult content is the second option [33]. A range of technical parental control tools can help parents to shield minors from age-inappropriate content [32], but the effectiveness of such solutions is debatable [6]. Unfortunately, many software solutions are ineffective because they either over- or under-block digital content [43]. These inaccuracies may be the result of guardians struggling with configuration options and user interfaces [7, 29]. Furthermore, Fleming *et al.* reported little difference in exposure to inappropriate material when blocking software was, or was not, installed [13]. A possible explanation is that minors are innovative enough to bypass such software. Moreover, there is also evidence that children resent their use, since they feel that the use of such software by their parents provides evidence of a lack of trust [28]. The reality is that the majority of guardians do not deploy automated filtering tools [31].

The previous two strategies are within the control of the guardian. When this is delegated to adult content providers, the deployed mechanism is usually either completely ineffective or overly intrusive. The ineffective approaches typically take the form of self-declaration, whereby individuals must declare their age or date of birth [52]. The process is elegant and effective at preventing *accidental* access. However, evidence suggests that many minors deliberately seek out, rather than accidentally stumble upon, adult content [13]. Livingstone *et al.* reported that 22% of surveyed 9-19 year-olds had purposefully accessed age-inappropriate content [25].

Some vendors go a step further, requiring a valid credit card, for example. This is based on the assumption that only adults can own a credit card but actually possession of such a token does not constitute proof of adulthood [39] especially since sufficiently motivated minors can easily obtain one [21].

The token-based technique crosses the invisible privacy boundary. Once privacy is violated, identification is carried out and age is established as a side effect. Vendors are able to gather far more information than is strictly required. Many other mechanisms, such as biometrics, also identify to authenticate. Biometrics seems an attractive alternative at first glance since pre-adolescent minors are smaller than adults, by and large, and this can easily be measured, especially as modern devices now routinely incorporate the requisite sensors. Unfortunately, the efficacy of the biometric-based approach has been questioned [49, 40]. Furthermore, there are genuine privacy concerns surrounding the collection and storage of biometrics belonging to minors without proper legislation mediating their use, and the storage and retention of such data [54, 14].

Another even more invasive solution is that of verifying identity by cross-referencing customer information such as name and address, date of birth with information held by other vendors or governments. This assumes that the other systems' data is accurate [39, 49] but there is evidence that such systems are easily manipulated by adolescents [48].

Based on this summary, we can conclude that the current range of third party mechanisms is unsatisfactory. They either present no bar to entry, or identify in order to establish adulthood, violating privacy. Moreover, many of these mechanisms introduce a substantial time lag into the access process which punishes adults instead of merely excluding children. The aim is to find a middle ground, something that estimates adulthood more effectively without identification. Such a solution would preserve the privacy of adult, or child, wanting to access adult content.

Using Images to Test Memory

There are two reasons for using images instead of simple questions to test knowledge. The first is that we inhabit the era of information being at everyone's fingertips, via search engines. If you ask someone a textual question they can search for the answer online — proving nothing more than an ability to search. If we use images to spark a memory it becomes harder for people to find the answer online.

The second reason is that people remember images better than the words than name them [27, 34, 35, 46]. Vision is the dominant human sense, being attended to before other modalities [37]. The sighted human effortlessly and almost subconsciously processes millions of images every day. Over time, an internal database of image forms and descriptors are accumulated, to support instantaneous recognition [46]. Lexical labels are linked with these objects, and every time the person interacts with the object and names it, that link gets stronger. Images appear to deliver two benefits: they are not as easy to search for, and they are more memorable than text.

We plan to use images in a recall-based context — not recognition as is the case with many psychological experiments proving the superiority of picture memory. Unlike recognition, recall requires provision of a lexical label from someone who is being cued by an image. There is evidence that people can label images they have seen before [8, 26]. Unfortunately, the labels are essentially proper nouns, and these are seldom descriptive, that can cause the memorial link between the name and the object to disappear or weaken over time [5]. There is also the "tip of the tongue" effect to consider [4, 18]. This, too, is particularly prevalent for proper noun labels. These labelling difficulties might impact the efficacy of our our proposed adulthood estimation mechanism, and we will return to this aspect in our discussion section.

Finally, it must be acknowledged that this mechanism excludes those with vision problems, and any deployment would have to be done in conjunction with some other mechanism to cater for those who cannot see the images well enough to label them.

DESIGN

The proposed mechanism acts to ensure that adults can access content but that minors are excluded. The concept is simple: an individual is presented with a grid of M images. They are required to label each image. The individual does not necessarily need to label all images correctly, but simply reach a threshold of N images to be classified as an adult. The expectation is that adults will be able to label at least N of M images but that minors will not. The mechanism does not effect identification, use biometric data or require the use of issued tokens, essentially preserving privacy.

The simplicity of this approach does not mean that achieving it is easy or simple. We faced two primary challenges in developing such a knowledge-based adulthood estimation mechanism: (1) finding the right set of images, and (2) delivering an accurate estimation.

(1) Image Identification

The first aim was to identify a set of candidate images for the proposed mechanism. The following *null* and *alternative* hypotheses were used to accept or reject a particular image:

$H1_0$: There is no difference between adults and minors when correctly identifying an image.

$H1_1$: There is a difference between adults and minors when correctly identifying an image.

(2) Estimating Adulthood

The challenge is that in estimating we inhabit the realms of probability. Hence we will make a judgement, and some of the time that judgement will be incorrect. We need to decide how many images to display, M, and how many of these the person should label correctly, N, to be judged an adult. The numbers chosen for N and M reflect the system's tolerance for error. The following hypotheses were tested:

$H2_0$: Adults will not have a different mean score than minors.

$H2_1$: Adults will have a higher mean score than minors.

$H2_2$: Adults will have a lower mean score than minors

The mechanism should judge adulthood to an acceptable level of accuracy. The classification of an individual can be categorised as a true positive, true negative, false positive or false negative. Table 1 illustrates the values of each, as a contingency table with the columns representing actual values and rows representing prediction values.

		Actual		
		Adults	*Minors*	
Prediction	*Adults*	True Positive	False Positive	*PPV*
	Minors	False Negative	True Negative	*NPV*
		Sensitivity	*Specificity*	*Accuracy*

Table 1. Contingency Table of Performance

Table 1 depicts the *sensitivity*, *specificity*, *positive predictive value*, *negative predictive value* and *accuracy*. The sensitivity (true positive rate) of the mechanism is the correct prediction of adults and the specificity (true negative rate) of the mechanism is its ability to exclude minors. The PPV, the positive predictive value, is the probability that the mechanism will correctly predict an adult. The NPV, the negative predictive value, is the probability that the mechanism correctly predicts a minor. PPV and NPV reflect the proportion of adults and minors in the sample and whether they are correctly assessed by the mechanism [1]. Lastly, accuracy is a measure of the true positives and true negatives in the population.

Delineating Children

Adults and children are generally defined in biological terms, with 18 years being the pivotal point as individuals enter adulthood on their 18th birthday. Different societies have different rules about people at various ages. In the UK people can marry at 16 and drink alcohol at 18, but in the US citizens have to be 18 to marry and 21 to drink alcohol. In some African countries marriage is allowed at 15, and in some European countries it is legal to drink alcohol at 16. Instead of focusing on 18 as a pivotal age, therefore, we should focus on the progression from childhood to adulthood, and the particular stage where they are most in need of protection.

The literature highlights the vulnerability of the 11-12 age group. This is when young people start spending less time with their families [23] and also start needing the support of adults outside the home [41]. If this is not provided at school they might well go online to get what they need. This is also the age when children, as a consequence of transferring from a relatively small primary school to a large secondary school, start to experience feelings of anonymity [44] and start interacting with the opposite sex [45] since they are embarking on puberty (earlier than in previous centuries, now to be expected at 11-12 [17]). Hence online anonymity is likely to appeal to them. At this age they start becoming more independent [3] and moving into adulthood and many are eager to embrace adult activities [20]. They need to be permitted increasing levels of autonomy to mature [10]. Parents and teachers therefore slowly relax controls to allow this. Unfortunately, the period of entering secondary school is when self-esteem decreases [53], and this only recovers when they settle properly into secondary school. Pre-adolescents are considered more vulnerable to abuse than older or younger children [12] and this reduced self-esteem might well be the reason for this vulnerability. Thus the 11-12 age group is a pivotal age, where efforts to enter the adult world need to be thwarted to protect them from that which they are not yet ready to deal with.

Finally, Fleming *et al.* [13] reported that younger adolescents were far less safety conscious when using the Web. Hence, in identifying images and testing our mechanism, we will focus on trying to distinguish between 11-12 year olds, and over 18s.

STUDY ONE

We started by identifying suitable images, then proceeded to test them in an estimation mechanism.

Identifying Images

An initial set of 104 images was gathered. Images were culled from various sources and aimed to include images of public figures and significant events, from the perspective of Western popular culture. The images within the set can be categorised, as follows: (1) Entertainment: music, film & television, (2) World Events, (3) Iconic Figures and (4) Politicians. Examples of images are: Michael Jackson, World Trade Centre, Dolly the Sheep and David Cameron.

Participants

60 minors and 74 adults participated. Participants aged 18 and over were recruited using traditional means, such as social networks and mailing lists. A number of schools were contacted in the United Kingdom, predominately Scotland and England, to obtain permission for their pupils to participate.

Procedure

The purpose of the survey was explained to the participant. The participant was then asked to complete a few simple questions pertaining to their nationality and age-range. Participants were then presented with each image and asked to label the image, i.e. identify the image content. The participants were asked to do this for each of the images in the initial set. The performance of each participant was analysed to determine if they had successfully identified the image or not. The label each participant submitted was analysed automatically using Levenshtein distance [16] to compensate for spelling errors.

Results

The survey was completed by all 134 participants, representing 60 minors and 74 adults. Adults scored higher on average (μ=54.32) than minors (μ=24.16) when labelling images within the initial set.

The next step was to identify those images that children struggle to label but that adults are able to label. The decision was taken to implement a threshold, as statistical analysis for all images was deemed unnecessary given certain images were easily identifiable by both groups. Consequently, images were selected when 75% more of adults correctly labelled it and only 25% or less of children correctly labelled it. The resulting 16 images were then statistically analysed using a Chi-Square test for association to determine if age-group was independent of performance. If performance for a particular image was not independent of age-group, it was deemed that $H1_0$ could be rejected and $H1_1$ accepted. In those cases were there were fewer than 5 participants in any of the contingency tables, used in the statistical analysis, a Fisher's exact test was used. The motivation for not using Fisher's exact test in all cases is that it may have been too conservative and lead to the rejection of viable images. Consequently, we deemed the Chi-square test adequate for the purpose of confirming image use and only defaulted to a Fisher's exact test when necessary.

The statistical analysis for each of the 16 images showed a strong association between adults and labelling performance when compared to minors. Consequently, in each case $H1_0$ was rejected and $H1_1$ was accepted and each image was added to the evaluation set.

Adulthood Estimation

The image identification stage pinpointed the image set for evaluation of the proposed mechanism. We could now proceed to test the mechanism. The image set of 16 images that emerged from the previous stage was used. The proposed mechanism was web-based and accessible from any device with a fully featured web-browser.

Participants

A total of 33 participants were recruited for the evaluation, comprising of 24 adults and 9 minors. Participants from the previous stage were excluded from this stage. A similar recruitment process to the previous stage was used.

Procedure

Participants accessed the system from a web browser. The first page provided information about the purpose of the investigation and what was expected. The participant was then asked for their actual age. The participant was than presented with six images and asked to label them. The system then predicted the user's age, based on the correctness of the image label, using Levenshtein distance [16] to compensate for spelling errors. The prediction was logged and participants thanked.

Results

The evaluation was completed by 33 participants, comprising 24 adults and 9 minors. Table 2 outlines the performance of the mechanism. The sensitivity of the mechanism, i.e. correct prediction of adults, is approximately 96%. The specificity of the mechanism, i.e. correct prediction of minors, is approximately 89%.

		Actual		
		Adults	*Minors*	
Prediction	*Adults*	23	1	*PPV* **≈96%**
	Minors	1	8	*NPV* **≈89%**
		Sensitivity **≈96%**	*Specificity* **≈89%**	*Accuracy* **≈92%**

Table 2. Mechanism Performance

An independent samples t-test was performed to determine if there was any difference in scores between adults and minors. Upon inspection of box-plots there were no outliers greater than 3 box-lengths from the edge of the box, but adult scores were not normally distributed, as assessed using a Shapiro-Wilk's test ($p < .05$). However, scores for minors were normally distributed, as assessed using a Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances for scores for adult and children, as assessed by Levene's test for equality of variances ($p = .870$). It thus became clear that a nonparametric test such as the Mann-Whitney U-test would be unsuitable due to the structure of the data, and the assumptions that this test makes. Consequently, the decision was taken to continue with the independent samples t-test. The reality is that adults scored higher (5.04 ± 1.197) than minors (1.78 ± 1.202). There was a statistically significant difference in mean scores between adults and minors, with adults scoring 3.26 (95% CI, 2.31 to 4.22) higher than minors, $t(31) = 6.968$, $p = < .001$.

The statistical analysis supported acceptance of H21 and rejection of $H2_0$ and $H2_2$. However, the small samples, unbalanced groups and the fact that one of the groups was not normally distributed, undermines the validity of our statistical analysis. Consequently, the data and analysis does not allow us to confirm or reject the hypotheses with any degree of certainty.

The aforementioned issues in the analysis could arguably be addressed using an alternative approach. A nonparametric equivalent to an independent t-test might well have produced stronger results. In this case, an acceptable alternative would have been a Mann Whitney U-test. However, nonparametric tests make assumptions about the structure of data. The Mann Whitney U-test, for example, requires the groups to be similarly structured. Given that this is not the case, we could only reasonably compare and report mean ranks rather than the scores themselves. Consequently, it was felt that adopting such an alternative approach would only serve to muddy the issue and limit subsequent discussions, as the focus would not be on mean scores but on ranks.

Discussion

The first study taught us lessons which fed into the next stage.

Image Identification: (1) Images of events appear to be unsuitable but images related to people can deliver the required distinctions if chosen carefully. (2) The best images for supporting adulthood-estimation are those that peak in the public media 15-20 years before the current date, and are then revisited from time to time so that they re-enter public consciousness.

Accuracy: (1) We should not ask participants to self-declare their age. In retrospect this was naïve. We ought to have two different sites for adults and children to use so that we are sure of the adulthood of the participant. (2) The first prototype used a challenge set comprising M=6 and N=2 to support a judgement. This might well set the bar too low.

STUDY TWO

Having learnt lessons from the first study, we commenced with our main study. We started with an image identification phase to identify a larger number of suitable images, then used these to trial a knowledge-based adulthood estimation prototype.

Image Identification

We expanded image topics beyond well-known people to also include images of vintage cars, old sweets and movies & television shows. We did not use images of widely publicised world events as the first study had shown these kinds of images to be suboptimal. We gathered a total of 481 images in categories: Famous People, Sweets, Cars and Film & Television Shows. The sweet name was obscured. It was, unfortunately, quite difficult to find suitable images of yesteryear's sweet wrappers. Car images depicted vintage cars with the logos and model names obscured.

The images used in the survey focused strongly on Western popular culture because we were exploiting people's exposure to images in public media and these are country-dependent, and we were going to test it primarily in the UK and the US. The main condition for an image to be considered suitable for use was whether adults, and not children, were likely to have been exposed to the image content (i.e. part of popular culture at least a decade ago).

Participants

As argued earlier, children under 13 years of age are considered to be most vulnerable. For this reason the image identification was carried out amongst S1 and S2 pupils (11 to 13 years of age) in two Scottish high schools. There was a total of 1601 participants, 1382 minors and 219 adults. A similar recruitment process was used as the previous study, with adults recruited using social networking services and mailing lists. Children were recruited by contacting various schools and obtaining permission from teachers.

Procedure

The first study asked participants to label all images. We could not realistically ask people to do this for all 481 candidate images. Moreover, the schools we contacted indicated that they could not allocate more than 5-10 minutes to this task, being under some pressure to complete their curriculum. We therefore decided to choose 30 images randomly to show to each participant. We did the same for adults and children.

We explained what the study was about, then collected demographics. We displayed 30 randomly chosen images, one at a time, and asked them to provide labels. Participants could also click on a button to pass. No cues were provided. Auto-fill in the label field was disabled. We recorded how long they took to label the images. We concluded by thanking them for their time and participation.

The system then determined the correctness of the labels. Participants' answers were compared to the correct labels for each image (both converted to lowercase, with punctuation and spaces removed). In case of the images of famous people, participants were expected to provide the name (eg. John Kennedy), not the role (eg. "President"), nicknames (eg. "JFK") or fictional character names in the case of actors (eg."Rocky"). A Levenshtein string distance algorithm [16] was used to accommodate spelling mistakes.

Results

1382 children and 219 adults participated in the image identification. The majority of child participants were born (1094) and raised (1114) in the UK.

When we examined the image labels it became clear that sweets and cars were unsuitable. Cars were too difficult even for adults to label. The Famous People and Film & TV categories delivered far better results since there was a clear divide between adults and children in terms of ability to label them. The mean score for children was 10% and for adults 26%.

Outcome

We realised that a timeout was not feasible when we considered the average labelling time. Web searching is so fast that we could not set any meaningful timeout period which would not also prevent people from typing in the answers they actually knew.

Since we were not able to test labelling of all images by all participants we had variable numbers of labels for our images, which barred reliable statistical testing. We thus decided that in order for an image to be considered suitable for use in the adulthood estimator it had to be labelled correctly by at least 50% of adults and by no more than 20% of children. This seemed to divide adults and children satisfactorily. A total of 44 images (9%) emerged from this stage.

Adulthood Estimation

To prepare the images we attempted to foil circumvention attempts, in this case searching for the image online. Since we could not feasibly prevent searching by implementing a timeout we decided rather to distort the images. The idea was to edit each image so that its content would still be recognisable to humans but not to software such as Google Image Search. Each image was edited using age-effect filter(old photography) as well as two artistic filters (ink outline and coloured pencil). However, for a few images this was not enough to deter an online image search. In this case, the images were edited using a texture filter (fissured) and an artistic filter (coloured pencil) which gave a very similar effect to that of age-effect filters combined with ink outlines and coloured pencil yet this indeed prevented an online image Search from returning any results. We disabled the right click and gave the images anonymous names.

Figure 1. Two faces, clear and distorted[50, 55]

Participants

For the purposes of the study three Scottish high schools were approached and asked to allow their first and second year pupils to participate in the survey. Adult participants were friends of those involved in the project but the majority of participants consisted of students from the University of Glasgow, both local and international. An invitation was also posted online.

Procedure

The experiment started with a page explaining what the study was about. Participants were instructed to label the images and to leave the label field blank if they could not. Demographics were collected, depending on whether this was the adult or child website.

Six *distorted* images were first displayed for labelling. The image set was randomised for each participant and no cues were provided.

One image in each set was a decoy image. Since many of our participants were children we did not want them to feel too discouraged by the process so the decoy image was one that they had a high probability of recognising. The correctness of the decoy image label was not used to judge adulthood. Depending on their performance, one of two things would happen:

- **Able to Make a Judgment:**
  If they labelled 4 or 5 images correctly it was likely that this was an adult (a Levenshtein string distance algorithm accommodated spelling errors.). If they got fewer than three correct (1 or 2), it was likely that this was a child. The next screen thus showed Challenge Set A again, *in the clear*, with the label fields prepopulated with the previously provided labels. They could then label the images again or re-submit initial labels.

- **Ambiguity:**
  If the system couldn't make a clear judgement a new set of images (Set B) was displayed. After they had labelled the new images, the first and second sets were displayed again, *in the clear*, in case this helped to jog their memory. They could change their initial labels.

The system concluded by telling the participant whether they had been judged adult or child, and thanked them for participating. A comment box was provided.

The reasoning behind this implementation was as follows. If we displayed 5 images then, based on the image identification results, we could expect adults to be able to guess 50% of these. This converts to ±3 images in the challenge set. If they identified four out of the five, they scored 80% and were classified as adult. If they identified only three, another set would be displayed. If they were able to identify three again (scoring 6 out of 9 = 67%), they would have passed the test and be classified as adult.

Children, based on our previous findings, only have a 1 in 5 chance of identifying images. If we presented them with 5 images we could expect them only to identify one or two of these. If by some fluke they managed to get three right in the first set, they would probably not be able to repeat that in the second set of images, and would be classified as a child.

Results

869 child participants aged 11 to 13 years participated with 356 adults participating. 79% of the adults were either British (31%) or American (48%). Child participants were mostly born (844) and raised (860) in the UK.

Accuracy:
Each participant was required to label at least 67% of the images correctly to be classified as adult. 99.7% of children were classified correctly when labeling distorted images and 99.08% when labeling in-the-clear images. At the same time, almost 34% of adult participants were classified correctly by the system when labeling distorted images and just over 43% when labeling the in-the-clear images.

		Actual		
		Adults	*Minors*	
Prediction	*Adults*	154	8	*PPV* **≈95%**
	Minors	202	861	*NPV* **≈81%**
		Sensitivity **≈43%**	*Specificity* **≈99%**	*Accuracy* **≈83%**

Table 3. Mechanism Performance for Non-Distorted Images

The poor results with respect to the adults were unexpected so we posted a random sample of 20 images from our image set on CrowdFlower, a crowd sourcing website. We posted two jobs: (1) one with the images distorted and (2) one with the images in the clear. Each job collected responses from 100 participants. We had to remove three fraudulent responses from the first job, and four from the second[1], but we were left with a large enough number of reasonable attempts to draw conclusions about recognition rates. The recognition rate for the distorted images was 38.66% whereas the recognition rate with the clear images was 49.31%. These numbers are remarkably close to the percentages in our study (33.89% and 43.26%), given the differences between the two population groups. This gave us more confidence in the paucity of our adult responses, suggesting that this was not an anomaly caused by our sample.

Adults scored better than children with the distorted images (2.14 ± 1.70) vs (0.18 ± 0.02). The assumption of homogeneity was violated, as assessed by Levene's test for equality of variances ($p < .001$). Therefore a Welch t-test is reported. There was a statistically significant difference in mean scores between adults and children, with adults scoring higher than children, 1.91 (95% CI, 1.71 to 2.10), $t(325.119)=19.29$, $p< .001$.

With respect to the clear images, adults scored higher (2.59 ± 1.80) than children (0.39 ± 0.73). The assumption of homogeneity was violated, as assessed by Levene's test for equality of variances ($p < .001$). Therefore a Welch t-test is reported. There was a statistically significant difference in the scores between adults and children, with adults scoring higher than children, 2.20 (95% CI, 1.99 to 2.41), $t\ (340.167) = 20.75$, $p < .001$.

Timings:
We examined the timings of responses and discovered a significant difference between the times taken by children and adults labelling the distorted images ($p < 0.001$).

[1]These were nonsense responses, with labels such as x entered for all images

There was no significant difference between the times adults and children took to label the clear images, however. We will discuss the implications of this in the next section.

DISCUSSION

Accuracy

Our knowledge-based adult estimation mechanism detected 99% of child participants, using N/M being greater than 67%. Unfortunately, the mechanism also classified too many adults as children (false negatives). An analysis was subsequently performed to investigate how the false positive and negative rate would change if the minimum score, N, was reduced (shown in Table 4). The percentage values for both child and adult participants represent the proportion of each group classified as adult.

	% False Positives		% True Positives	
Min Threshold	Distorted	Clear	Distorted	Clear
20% (1 out of 5)	17.91	30.15	82.22	86.24
40% (2 out of 5)	4.45	9.2	64.44	72.47
44% (4 out of 9)	0.51	2.3	41.67	50.84
56% (5 out of 9)	0.3	1.61	38.06	47.75
60% (3 out of 5)	0.8	2.88	43.33	51.97
67% (6 out of 9)	0.3	0.92	33.89	43.26

Table 4. Tweaking Thresholds

The distortions might have deterred recognition. Some comments referred to this: *"It was really, really difficult to see who they were when the pictures werent clear"*. If the distortions made recognition difficult, it is indicative of the unfortunate yet inevitable tension between preventing circumvention and maximising accuracy. Still, this would not have made the adults fail the test, since we displayed the images in the clear too.

Adults took significantly longer to label the distorted images. What were they doing differently? Four different things might have happened when they examined the image: (1) they recognised the person, and knew the label; (2) they recognised the person but couldn't come up with label; (3) the distortions made it difficult to be sure; or (4) they did not know who the person was.

(1) does not explain the time lag. Using an online image search would be the natural thing to do if (2) were true. With no timeout this strategy might have been deployed. One of the comments confirms that this happened at least in one case *"I admit that I googled Lisa Kudrows last name, because I knew who she was but just could not remember the last name for whatever reason."*.

Was this a manifestation of tip-of-the-tongue syndrome [4]? Eight participant comments alluded to this eg. *"I think recognized the guy from Monty Python (first picture) but couldnt remember his name (not Michael Palin) and felt I shouldnt look it up."*.

If (3) were the case the time lag could simply be due to a participant spending some time examining the picture. There were three comments related to this, including: *"thank goodness there were clearer pictures later!"*. In this case circumvention resistance clearly impacted accuracy.

What about (4)? We used images that 50% or more of adults were able to label, but that is a probability, not a certainty. It could have been the case that the test was just too hard. They could have attempted a online image search, which would explain the time lag.

We believe that the images presented the adults with too much of a challenge. Those who knew enough about the depicted person to search probably found the name via a search engine. Some may have felt that this was cheating, but it is fairly safe to conclude that the combination of the hard to identify images, and the requirement for a pass rate of over 67%, probably led to the estimation mechanism delivering too many false negatives.

It might be reasonable to lower the minimum threshold to 40% (2 out of 5) or to 60% (3 out of 5). On the other hand, as more adults are correctly classified, so the number of false positives, in terms of children being classified as adults, increases (see Table 4). This is where context comes in. Some adult content is far more sensitive than other, as reflected by the different ratings applied to films, and the tolerance for error might well be aligned to the nature of the content being protected by the access control mechanism.

Foiling Circumvention Attempts

We used distorted images which would deliver no results if fed into image search engines. So we know that an online image search would not have delivered any results. Any adulthood-estimation mechanism has to contend with the ingenuous efforts of determined children. What would an enterprising child do if he or she were determined to access adult content and this mechanism was barring entry? The child could feasibly post the image onto a social networking site, asking others to identify it. This, of course, would take time and even a relatively long timeout could help to foil this tactic. Such a posting might also alert a responsible adult to the child's activities. They could also send the image to an older friend or sibling and ask for help in identifying it. When we rejected the idea of a timeout earlier that was because we couldn't prevent online searching with a timeout. It does seem more feasible to prevent children consulting other children with a timeout since that will probably take longer.

We should also implement some kind of mechanism to discourage repeat attempts. A doubling delay would be very effective — so that no further requests are accepted from a particular IP address for a certain period if the previous attempt failed.

Future Work

The primary challenge is addressing deployment-related issues. **The first** issue is bootstrapping the solution with a viable image set. Whereas a time-consuming collection and curating process was reasonable to support a research project, it is unrealistic in a real-life solution. Consequently, an automated process to collect and curate images is required. **Secondly,** such a process would have to incorporate some notion of periodic refreshing of the image set to tackle evolving public consciousness and improvements in online search tools. Furthermore, minors will adopt other strategies to defeat the mechanism, eg. creating online communities to exchange answers. Moreover, even if an automated process could rapidly deliver viable images, it would encounter difficulties related to the number of well-known personalities available for use. There is a finite number of individuals and events that can be used in this context.

Nevertheless, even if all the aforementioned issues are tackled and an evergreen image set *can* be created, **the convenience** of such a solution is debatable. The solution may seem viable when presented prior to consuming a lengthy 2-hour motion picture, but seems excessive if the user is consuming several short videos via YouTube, for example. These challenges would have to be overcome if this mechanism's potential is to be realised in practice.

If we are able to resolve the deployment challenges, two aspects of the current solution require attention. **(1) Setting N and M:** What should N and M be? This is likely to depend on the level of certainty required that the user is an adult. The required certainty will probably be linked to the content it is shielding, and we will consider this a matter for future research.

(2) Easing The Process for Adults: To ameliorate the tip of the tongue problem we could accept meta-data instead of a label. This would lead to greater complexity since even more spelling errors would have to be accommodated and it would slow the estimation process down. Moreover, search engines already help people to fix on a name given meta data, development in this area does not seem warranted.

CONCLUSION

In this paper we presented a knowledge-based mechanism for classifying an online user as either an adult or a child. The mechanism was able to correctly classify the a user as a child with 99% accuracy. It also currently delivers an unacceptably high rate of false negatives, barring far too many adults. It is a clear improvement over self-declaration, and, most importantly, does not violate the user's privacy.

It would be naïve to believe that we can craft a mechanism that uses one technique, in solo, to solve this tricky problem of adulthood estimation. Our desire for simplicity of purpose and solution cannot be satisfied in this case. There are some directions to take this research, which are certainly worth pursuing, since parents currently face unprecedented challenges in protecting their children.

Acknowledgements

We acknowledge the essential contributions of Charlene Low and Aga Kielbinska, who worked on the first and second phases of this project, respectively. Without their sterling work we could not have completed this project. We also thank our many participants and the teachers who made the evaluations possible.

REFERENCES

1. D. G. Altman and J. M. Bland. Diagnostic tests 2: predictive values. *BMJ: British Medical Journal*, 309(6947):102, 1994.
2. BBC3. BBC three survey reveals one in four young people first view porn at age 12 or under. `http://www.bbc.co.uk/mediacentre/latestnews/2014/porn-whats-the-harm.html`.
3. T. J. Berndt. The features and effects of friendship in early adolescence. *Child Development. Early Adolescence (Dec., 1982)*, 53(6):1447–1460, 1982.
4. R. Brown and D. McNeill. The "tip of the tongue" phenomenon. *Journal of verbal learning and verbal behavior*, 5(4):325–337, 1966.
5. D. M. Burke, J. K. Locantore, A. A. Austin, and B. Chae. Cherry pit primes brad pitt homophone priming effects on young and older adults' production of proper names. *Psychological Science*, 15(3):164–170, 2004.
6. M. Buzzi. Children and YouTube: Access to Safe content. In F. Cipolla-Ficarra, K. Veltman, D. Verber, M. Cipolla-Ficaeea, and F. Kammüller, editors, *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, CHItaly, pages 125–131, New York, NY, USA, 2011. ACM.
7. M. Buzzi. What are your children watching on YouTube? In *Advances in New Technologies, Interactive Interfaces and Communicability*, pages 243–252. Springer, 2012.
8. C. B. Cave. Very long-lasting priming in picture naming. *Psychological Science*, 8(4):322–325, 1997.
9. D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Commun. ACM*, 28(10):1030–1044, Oct. 1985.
10. V. Chirkova, R. M. Ryan, Y. Kimc, and U. Kapland. Differentiating autonomy from individualism and independence: A self-determination theory perspective on internalization of cultural orientations and well-being. *Journal of Personality and Social Psychology*, 84(1):97–110, 2003.
11. J. Edgar and G. Rayner. Judge blames easy access to online pornography after boy, 12, rapes seven-year-old sister, 4 March 2014. The Telegraph.
12. D. Finkelhor and L. Baron. High-risk children. In D. Finkelhor and Associates, editors, *Sourcebook on child sexual abuse*. Sage, Newbury Park CA, 1986.
13. M. J. Fleming, S. Greentree, D. Cocotti-Muller, K. A. Elias, and S. Morrison. Safety in cyberspace adolescents' safety and exposure online. *Youth & Society*, 38(2):135–154, 2006.
14. A. M. Froomkin. The death of privacy? *Stanford Law Review*, pages 1461–1543, 2000.
15. F. Gilbert. Age verification as a shield for minors on the internet: A quixotic search? *Shidler Journal of Law Commerce & Technology*, 6, 2008.
16. M. Gilleland. Levenshtein distance, in three flavors. *Merriam Park Software: http://www. merriampark. com/ld.htm*, 2009.
17. P. D. Gluckman and M. A. Hanson. Evolution, development and timing of puberty. *TRENDS in Endocrinology and Metabolism*, 17(1):7–12, 2006.

18. T. H. Gollan, R. I. Montoya, and M. P. Bonanni. Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology*, 19(3):278, 2005.
19. M. Harbach, S. Fahl, M. Rieger, and M. Smith. On the acceptance of privacy-preserving authentication technology: The curious case of national identity cards. In E. De Cristofaro and M. Wright, editors, *Privacy Enhancing Technologies*, pages 245–264. Springer, 2013.
20. D. F. Jacobs. Juvenile Gambling in North America: An Analysis of Long Term Trends and Future Prospects. *Journal of Gambling Studies*, 16(2-3):119–152, 2000.
21. H. Kanda, Y. Osaki, T. Ohida, Y. Kaneita, and T. Munezawa. Age verification cards fail to fully prevent minors from accessing tobacco products. *Tobacco control*, 20(2):163–165, 2011.
22. D. Kirby. Sex and HIV/AIDS education in schools. *BMJ: British Medical Journal*, 311(7002):403, 1995.
23. R. Larson and M. H. Richards. Daily companionship in late childhood and early adolescence: Changing developmental contexts. *Child Development*, 62(2):284–300, 1991.
24. P. W. Linville. The complexity–extremity effect and age-based stereotyping. *Journal of personality and social psychology*, 42(2):193, 1982.
25. S. Livingstone, M. Bober, and E. Helsper. Internet literacy among children and young people: findings from the UK children, 2005. Go Online project `http://eprints.lse.ac.uk/archive/00000297(Accessed04/06/2010)`.
26. C. M. MacLeod. Forgotten but not gone: Savings for pictures and words in long-term memory. *Journal of Experimental Psychology*, 14(2):195–212, 1988.
27. S. Madigan. Picture memory. In J. Yuille, editor, *Imagery, memory, and cognition: essays in honor of Allan Paivio*, pages 65–89. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.
28. G. Mascheroni. Parenting the mobile internet in italian households: Parents' and children's discourses. *Journal of Children and Media*, (Published online: 02 Sep 2013):1–17, 2013.
29. L. A. McFarlin, K. J. Buffardi, and R. M. Schumacher. Usability impact on effectiveness of parental controls. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 51, pages 1039–1043. Sage Publications, 2007.
30. F. Mishna, C. Cook, M. Saini, M.-J. Wu, and R. MacFadden. Interventions for children, youth, and parents to prevent and reduce cyber. *Campbell Systematic Reviews*, 2, 2009.
31. K. J. Mitchell, D. Finkelhor, and J. Wolak. Protecting youth online: Family use of filtering and blocking software. *Child abuse & neglect*, 29(7):753–765, 2005.
32. H. F. Odabaşi. Parents views on internet use. *Turkish Online Journal of Educational Technology*, 4(1), 2005.
33. Ofcom. Ofcom report on internet safety measures. strategies of parental protection for children online, 15 January 2014. `http://stakeholders.ofcom.org.uk/binaries/internet/internet-safety-measures.pdf`.
34. A. Paivio. *Imagery and verbal processes*. Holt, Rinheart & Winston, New York, 1971.
35. A. Paivio, T. Rogers, and P. Smythe. Why are pictures easier to recall than words? *Psychonomic Science*, 11(4):137–138, 1968.
36. J. Peter and P. M. Valkenburg. Adolescents' exposure to sexually explicit Internet material and sexual satisfaction: A longitudinal study. *Human Communication Research*, 35(2):171–194, 2009.
37. M. I. Posner, M. J. Nissen, and R. M. Klein. Visual dominance: an information-processing account of its origins and significance. *Psychological review*, 83(2):157, 1976.
38. C. J. Rathbone, C. J. Moulin, and M. A. Conway. Self-centered memories: The reminiscence bump and the self. *Memory & Cognition*, 36(8):1403–1414, 2008.
39. B. Reznikov. "Can I See Some ID?" Age Verification Requirements for the Online Liquor Store. *Shidler Journal of Law, Commerce & Technology.*, 4:5–13, 2007.
40. C. Roberts. Biometric attack vectors and defences. *Computers & Security*, 26(1):14–25, 2007.
41. R. W. Roeser, C. Midgley, and R. C. Urdan. Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Pscychology*, 88(3):408–422, 1996.
42. D. C. Rubin. *Autobiographical memory across the lifespan*. MIT Press Cambridge, MA, 2002.
43. K. G. Schneider. *A practical guide to Internet filters*. Neal-Schuman Publishers, Inc., 1997.
44. R. G. Simmons and D. A. Blyth. *Moving into Adolescence: The Impact of Pubertal Change and School Context*. AldineTransaction, 2009.
45. R. G. Simmons, R. Burgeson, S. Carlton-Ford, and D. A. Blyth. The impact of cumulative change in early adolescence. *Child Development*, 58(5):1220–1234, 1987.
46. S. N. Singh, M. L. Rothschild, and G. A. C. Jr. Recognition versus recall as measures of television commercial forgetting. *Journal of Marketing Research*, 25(1):72–80, 1988.
47. A. Spence. 200,000 children are watching hardcore porn, 28 March 2014. The Times. `http://www.thetimes.co.uk/tto/news/medianews/article4047202.ece`.
48. B. Szoka. COPPA 2.0: The New Battle over Privacy, Age Verification, Online Safety & Free Speech, 24 May 2009. http://techliberation.com/2009/05/24/coppa-20-the-new-battle-over-privacy-age-verification-online-safety-free-speech/.
49. A. Thierer. Social networking and age verification: Many hard questions; no easy solutions. *Progress on point*, 14:1–33, 2007.
50. U.S. National Archives and Records Administration. Richard Nixon. `http://upload.wikimedia.org/wikipedia/commons/3/39/Richard_M._Nixon,_ca._1935_-_1982_-_NARA_-_530679.jpg`.
51. L. Wade. Facing the threat: Invading the body for national security. *Knowledge, Technology & Policy*, 17(1):74–80, 2004.
52. L. G. Walters. Method of age verification for electronic media. US Patent 6959861, 2003.
53. A. Wigfield, J. S. Eccles, D. MacIver, D. A. Reuman, and C. M. Midgley. Transitions during early adolescence: Changes in children's domain specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Pscyhology*, 27(4):552–565, 1991.
54. J. D. Woodward, N. M. Orlans, and P. T. Higgins. *Biometrics:identity assurance in the information age*. McGraw-Hill/Osborne New York, 2003.
55. World Economic Forum. Tony Blair. `https://www.flickr.com/photos/worldeconomicforum/3488873590/in/photostream/`.