ボット判定の機会を活用した情報セキュリティ・倫理の学習手法の知 識習得と知識持続効果の検証

吉川 諒^{†a)} 落合 秀也^{†b)} 矢谷 浩司^{†c)}

Examining the Effects on Knowledge Acquisition and Retention about Information Security and Ethics through Opportunistic Learning in Human-verification Tasks

Ryo YOSHIKAWA^{†a)}, Hideya OCHIAI^{†b)}, and Koji YATANI^{†c)}

あらまし 情報セキュリティ・倫理に関する知識を学ぶ必要性がますます高まる中,全てのユーザが最新の知識を定期的に手に入れられているとは言い難い。そこで我々は先行研究において,一般人に対して定期的に学習機会を提供するためのシステムの設計を目指した。そして,情報倫理・セキュリティに関するマイクロラーニングをボット判定に統合するシステム,DualCheck を提案した。DualCheck は reCAPTCHA-v2 等と同様にチェックボックスをクリックする挙動でボット判定を実施するが,同時にインターネットの適切な利用を促すクイズを出題する。先行研究では DualCheck を利用することが知識の定着に役立ち,またユーザビリティが高いことが示されたが,定着した知識がどの程度持続するかは不明であった。そこで本研究では,DualCheck 利用後4ヶ月程度が経過した参加者を対象に,知識の定着を測定するテストを行った。結果として,DualCheck の利用前に比べて,テストの正答率が有意に高いことが明らかとなった。更に,DualCheck 利用直後の知識の多くが持続していることも判明した。

キーワード マイクロラーニング,情報セキュリティ教育, CAPTCHA,情報倫理

1. はじめに

近年多くのユーザがインターネットのサービスやコミュニケーションを日常的に利用するようになり、インターネットを利用する上での倫理やリスクを理解することは不可欠である。しかし、一般のユーザが情報倫理・セキュリティを学ぶ機会が十分に確保されているとは言い難い。情報処理推進機構 (IPA) の調査 [1] によれば、スマートデバイス利用者のうち、情報倫理に関する教育を受けたことがあると回答した人はわずか17.9%であった。さらに、そのようなトレーニングは職場や学校で行われることが多く、これらに属さない人にとって特に機会が不足していると考えられる。

これらの問題を解決するためには, インターネット ユーザが日常的にこれらの知識を得られる仕組みが重 要である. 気軽に学習できる仕組みにするためには、 インターネット利用者が既に慣れ親しんでいる既存の インタラクションを活用することが望ましい. そこで 筆者らは、先行研究 [2] において DualCheck という、 ウェブアプリケーションのプロトタイプを提案した [2]. DualCheck は、アンケートフォームなどでボット を排除するための CAPTCHA [3] のようなボット判定 機会を活用する学習支援システムである. DualCheck の挙動を図1に示す. DualCheck は、従来のボット判 定タスクに代わって情報セキュリティ・倫理に関する 問題を出題するシステムであり、ユーザは問題に回答 することで, 同時にボット判定を通過することができ る. ユーザは、DualCheck が出題した問題に対し、5 つの選択肢のうち1つをクリックして回答する.シ ステムは、正答と解説を提示して、ユーザの回答から 5 秒後に次のページに移動するボタンを有効にする. DualCheck のボット判定は、ユーザが問題に正答した か否かに依存しない、ボット判定は、ユーザが選択肢 を選ぶ際の挙動をもとに、reCAPTCHA-v2 などと同 様の方法でボット判定を行うことが期待される. この

[†]東京大学

The University of Tokyo

a) E-mail: ryo@iis-lab.org

b) E-mail: ochiai@elab.ic.i.u-tokyo.ac.jp

c) E-mail: koji@iis-lab.org

DOI:10.14923/transj.?????????



図 1 DualCheck のインターフェイス. 左:オンラインフォームに組み込まれた DualCheck の例. なお,本研究は学習効果を検証することを目的としているため、現プロトタイプにはボット判定の機能は組み込まれていない. 中央: DualCheck が提示する問題を読んで、ユーザが選択肢の一つを選ぶ. 現在は reCAPTCHA-v2 のインタフェースの挙動を模倣している. 右:DualCheck は、与えられた問題に対する正解と解説を提示する. 正解と解説が表示されてから 5 秒後に送信ボタンが有効になる. この時間を利用してユーザは正解・解説を読むことができる. 本システムでは、ユーザが正解を選択したか否かはボット判定の結果に影響せず、reCAPTCHA-v2 のようにクイズに回答する際の挙動で判定することを想定している. / The DualCheck interface. Left: DualCheck can be integrated into online forms as a human verification mechanism. We note that our current prototype does not implement the human verification mechanism because our primary objective of this work is to validate the effect of opportunistic microlearning through DualCheckinstead of evaluating the robustness of human verification. Middle: Users choose one of the choices after reading the question presented by DualCheck. The current implementation simply pretends to be performing human verification like reCAPTCHA-v2. Right: DualCheck presents the correct answer and explanation about the given question. The system enables the submit button 5 seconds after it shows the correct answer and explanation. In this manner, users have an opportunity to read them. The system does not consider whether users have chosen the correct answer or not for human verification. Instead, it is expected to perform human verification through a mechanism like reCAPTCHA-v2.

ように、DualCheck は、信頼性の高いボット判定を行うと同時に情報セキュリティ・倫理に関するマイクロラーニングの機会を提供する.

先述の先行研究では、15日間のユーザ実験を行い、DualCheck を用いることによる知識定着効果と DualCheck のユーザビリティの評価を行った。結果として、検証に用いた 10 間のうち 9 間でユーザ実験前後での正答率の有意な向上が確認された。また、これらの類題 10 間のうち 5 間において、DualCheck を利用していない一般的なインターネット利用者と比較して有意に高い正答率が確認された。また、DualCheck のユーザビリティは、テキストベースの CAPTCHA や画像ベースの reCAPTCHA よりも有意に高く評価された。定性的な結果からも、参加者の DualCheck に対する好意的な態度が確認された。

一方で先述の研究では、DualCheck を 2 週間利用し続けた翌日に情報セキュリティ・倫理に関する問題を出題して正答率を測定しているが、学習から時間が経過した際に知識がどの程度持続しているかの調査が行われていない。Reinheimer らの研究 [4] では、フィッシング対策の教育の効果が学習後 4 ヶ月程度を境に低下し始めたことが確認された。そこで本研究では、筆者らの過去の研究 [2] の参加者に対して、実験後 4 ヶ月頃を目処に再度テストを実施した。結果とし

て、DualCheck 利用前の参加者の正答率や、DualCheck 非利用者の平均正答率(一般正答率、6.2 節で定義)より平均正答率が有意に高いことが判明した。更に、DualCheck 利用直後のテストの平均正答率との有意な差は確認されず、DualCheck を介して得た知識の多くを維持していることが示唆された。

2. 関連研究

2.1 情報倫理・セキュリティ教育に関する研究

多くの人がパソコンやスマートデバイスを持ち日常的にインターネットを利用している現代において、インターネットの倫理や脅威に関する知識は欠かせない。しかし、それらの知識を学ぶ機会が少ないことから、十分な知識を有しておらずリスクを抱えている人が多い。2019 年に情報処理推進機構 (IPA) が行った調査 [1],[5] によると、情報倫理の教育を受けたと回答した 10 代は 38.0% にとどまる。この割合は高年齢層になるとさらに低くなり、例えば 70 代の調査ではわずか9.6% となっている。また、情報倫理教育を受けたことがあると回答した人のうち76.8% は、学校や職場で教育を受けたと述べている。さらに、多くの人が情報セキュリティに関する基礎的な知識を有していないと判明した [5]. 米国の Grimes らの研究でも同様の結果が得られており、高齢者は、大学生と比べてインター

ネットセキュリティの危険に対する知識や意識が低い ことが明らかになった[6]. これらの調査結果から、学 習機会が限られており、情報セキュリティや倫理に関 する知識の定着が十分でないことが示唆される.

これらを踏まえ、手軽にインターネットリテラシーを身につけるために、ゲームを活用する研究がなされてきた。Shengら [7] はアンチフィッシングの知識をビデオゲームに組み込んでいる。さらに、彼ら [8] は、ロールプレイに基づくフィッシング攻撃を行うことで、人々の知識の不十分さを認識させた。一方で Druryら [9] によれば、攻撃や脅威が進化し、より複雑で巧妙になっていることから、学習方法についてさらなる検討が必要である。また、一度教育を受けるだけでは不十分である。Reinheimerらは、フィッシングに関する啓発・教育プログラムの有効性を調査した [4]。彼らは、参加者のフィッシング識別能力は、教育プログラムの4ヶ月後までは維持されていたが、それ以降は低下することを発見した。この結果は、定期的なトレーニングが重要であることを示すものである。

我々が提案した DualCheck [2] は、情報セキュリティや倫理の学習のために、CAPTCHA を活用するシステムである。多くの人が、インターネット利用時にCAPTCHA に定期的に遭遇するため、既存の学校カリキュラムや職場におけるトレーニングよりも頻繁に学習機会を提供できる。本研究では、DualCheck を利用してから4ヶ月後に再度テストを行うことで、知識の持続効果を確認する。

2.2 マイクロラーニングに関する研究

マイクロラーニングとは、短時間で繰り返し学習す る方式のことである. マイクロラーニングのタスク は、学習者が短時間で完了できるように、小分けにし て設定されている. また, マイクロラーニングのもう 一つの利点は、ユーザのインタラクションやタスクに 組み込むことで、臨機応変に学習の機会を提供する ことができる点である. HCI の分野では、マイクロ ラーニングの学習効果について研究されてきた. 多く のプロジェクトは、マイクロラーニングシステムを通 して、語彙の習得をターゲットにしている. Trusty と Truong [10] は、Web ページ上の英単語を、ユーザが 学習中の外国語に自動的に翻訳するブラウザ拡張機能 を開発した. このシステムを用いることで、ユーザは ウェブページを読みながら臨機応変に語彙を学習する ことができる. Cai ら [11] は、テキストチャット中の ユーザの待ち時間を利用した語彙学習システムを構 築した. Dingler ら [12] は、モバイル通知を利用したマイクロラーニングのために QuickLearn を実装した. QuickLearn では、モバイル通知により単語の質問を提示し、ユーザが移動中や限られた時間しかない場合でも、簡単な語彙学習をすることが可能になった. これらの研究では全て、単語の学習効果が確認された.

マイクロラーニングを語彙学習に援用する研究が多くなされてきた一方で、情報セキュリティ・倫理教育にマイクロラーニングを活用する研究も行われた. Mohammed ら [13] は、小学生向けの ICT 教育にマイクロラーニングを取り入れた研究を行った. 彼らは、フラッシュカードやビデオを用いたマイクロラーニングは、教科書ベースの教育と比較して、学習能力を最大 18% 向上させ、長期記憶の保持を向上させることを発見した. この研究の調査は、情報セキュリティ・倫理に関するマイクロラーニングが CAPTCHA に組み込まれた場合にいかに効果的であるかを示唆しており、本研究では、実際に DualCheck を利用することで得た知識の持続効果を確認する.

2.3 CAPTCHA に関する研究

悪意のある自動アクセスを防ぐために、ユーザと ボットを判別する仕組みが多くのウェブサイトで導入 されている. CAPTCHA [3] は、Ahn らによって開発 されたボット判別システムで、最も広く使用されてい るものの一つである. 初期の CAPTCHA は、視覚的 に歪んだ文字列を正しく入力することをユーザに要求 していた. また, Ahn らは reCAPTCHA [14] も開発し た. reCAPTCHA はオリジナルの CAPTCHA と同じ 機能を提供するとともに、ユーザの回答を活用するこ とで OCR ソフトウェアの認識正答率改善にも貢献す る. しかし、これらの CAPTCHA は機能性や使い勝手 に問題があることも認識されている. Jeff ら [15] は、 テキストベースの CAPTCHA の頑健性とユーザビリ ティについて議論した. 彼らは、テキストの歪みの程 度や、紛らわしい文字の存在によるユーザビリティの 懸念を指摘している.

このようなユーザビリティの問題に対処するために、別の形式のボット判定タスクを検討する研究も行われている. Yamamoto ら [16] は、4 コマ漫画の並び替えを行うタスクを考案した. Fanelle らは、主に視覚障害者が使用する新しい音声 CAPTCHA を設計した [17]. 彼らの設計は、正答率と速度の点で既存の音声 CAPTCHA よりも優れていた. また、最近の CAPTCHA の開発により、ボット判定のインタラクションは負担の少

ないものとなり、例えば reCAPTCHA-v2 はチェックボックスをクリックするだけでよく、reCAPTCHA-v3 はユーザに操作を要求することもない。Tanthavechら [18] は、複数の CAPTCHA のユーザビリティを比較し、簡単な計算問題を解くようにユーザに要求する math CAPTCHA が、5つのボット判定タスクの中で最も高い使い心地の評価を得たことを示した。この理由として、このようなタスクが手軽な頭の体操になった可能性が考えられる。このように、ボット判定のタスクは、ユーザが直接的にメリットを感じることができれば、より受け入れられやすくなることが期待される。

そこで筆者らの先行研究 [2] では、CAPTCHA に有意義な役割を与えることがユーザビリティ改善にもつながると考え、情報倫理・セキュリティの知識提供を行う CAPTCHA を設計提案した.

3. DualCheck

3.1 システム実装

本節では、筆者らの先行研究[2]において設計した DualCheck の内容について説明する. DualCheck は, reCAPTCHA-v2 と同様に、チェックボックスをチェッ クするボット判定タスクを通じて、情報セキュリティ・ 倫理に関するマイクロラーニングの機会を提供するも のである. DualCheck は、情報セキュリティ・倫理に 関する選択肢式の問題をユーザに提示する. ユーザは このクイズに答えて知識を身に付けるのと同時に、選 択肢を選ぶ動作によってボット判定を突破することが できる. 図1は、Web上で実装したインタフェースで ある. このインタフェースは、従来の CAPTCHA のよ うにウェブフォームに容易に組み込むことができる. DualCheck は情報セキュリティ・倫理に関する2つの 小問の組み合わせを出題し、ユーザはそれぞれの文が 正しいかどうか判断する. 回答は選択肢式で、「A の み正しい」「B のみ正しい」「両方正しい」「両方誤り」 「わからない」の5つの選択肢が表示される. ユーザ がいずれかのチェックボックスにチェックを入れる と、DualCheck は正解と短い説明を提示し、適切な知 識の習得を促す. また, 正解が表示された5秒後に次 のページへ移動するボタンを表示するようにすること で,正解と解説を読むことを促している.

なお、本システムでのボット判定は、ユーザが問題に正答したか否かに依存しない。ボット判定は、reCAPTCHA v2 と同様に、チェックボックスをクリックしたときのカーソルの挙動を解析することで行わ

れることを想定している。また、本研究で開発したDualCheckのプロトタイプは、実際にはボット判定機能を有していない。これは、reCAPTCHAのコードが公開されていないためである。一方でシステムの挙動の見た目は reCAPTCHA-v2 を模しており、ユーザにボット判定機能の有無を伝えない限りユーザが機能の有無に気づくことは難しいと考えられるため、実際にボット判定機能を有していないことによる知識の定着効果や使いやすさへの影響は本実験においてはないものと考えられる。なお、第4章で述べるユーザ実験でも、DualCheckのプロトタイプがボット判定機能を有していないことは参加者に伝えていない。

3.2 問題の作成

筆者らの先行研究[2]では、自ら出題用の問題を作 成した. 問題の作成にあたり、2つの基準を設けた. 1つ目はインターネットの安全と倫理に関する基本的 な内容であること、2つ目は難しすぎず、知られすぎ ていないテーマであることである. 上記の条件を満た すために高校の指導要領の解説資料 [19] や、IPA の発 表した情報セキュリティ 10 大脅威 [20] などを参考に し、問題の草案を作成した、筆者のうちネットワーク セキュリティの専門家 1 名が、自明なものや文意が不 明確なものを除く作業を行い、25 問を選定した. そし て、クラウドソーシングを用いて、実際に参加者に問 題を解いてもらうことで問題の難易度が適切であるか, 文意が齟齬なく伝わるかを調査した. 調査の結果 25 問 中 15 問が 80% 以上の正答率となった. これらの質問 は、人々がすでに知識を有している内容であり、ユー ザ実験には適していないと考えた. そこで, 正答率が 80% 以下の10間を選び、細かな文言を修正してユー ザ実験に用いることとした. この 10 問のテーマには, SNS でのプライバシーの露出や誹謗中傷、フィッシン グに関するものなどに加え、cookie の役割や、USB メ モリ接続の際の注意点などが含まれる.

4. 15 日間に渡るユーザ実験

4.1 実験の概要

筆者らの先行研究 [2] では、DualCheck の評価のために 15 日間のユーザ実験を実施した. このユーザ実験は、DualCheck を利用することによる知識定着効果と使いやすさを評価するために行われた.

調査参加者には、実際の利用状況に近い方法で利用 してもらうためにデセプション実験を設計し、本来の 目的を知らせないで調査を行った. 調査は 15 日間に 渡り、まず初日に情報セキュリティ・倫理に関するクイズ(Q1-Q10)を含む問題を出題した。その後、2-14日目に、1日3回(14日目には4回)ウェブアンケートへの回答を依頼し、回答送信前に DualCheckで1問のクイズを出題した。このクイズは先述の Q1-Q10のなかから1問を選んで出題し、参加者は、この期間内にそれぞれのクイズを各4回解いた。15日目に本来の調査の目的を説明し、参加者全員の合意が得られたため調査を続行した。15日目には、Q1-Q10とその類題(Q1a-Q10a)を含む問題を出題するとともに、DualCheckの使い心地についてアンケートで質問した。

DualCheck の学習効果の比較のため、二つの出題 モードを用意し、それぞれ OneTime モード、Repeat モードと名づけた。OneTime モードは、参加者が選択 肢にチェックを入れた後に解説を表示し、選択肢の正 誤にかかわらず次に進めるものである。一方 Repeat モードは、ユーザの答えが誤っていた場合、解説を提 示した後に再び同じ問題を出題し、正解にチェックを 入れるまで強制的に問題を解かせるものである。正解 を選んだ後にフォームを送信するボタンが表示される ようにした。

参加者はクラウドソーシングサービスで募集を行い、34名の参加者(女性25名, 男性9名, 20代6名, 30代9名, 40代13名, 50代4名,60代2名)が集まった. 34名をランダムに分け、このうち16名にOneTimeモード、18名にRepeatモードを設定してDualCheckを利用してもらった。参加者全員には謝礼として税込2800円が支払われた.

4.2 調査の結果

先行調査 [2] では、学習効果とユーザビリティ評価の両面での成果が見られた、結果の概要を示す.

4.2.1 学習効果の検証

今回の調査の第一の目的は、マイクロラーニング に対する DualCheck の効果を検証することである. そこで、まず、調査で使用した 10 間の質問(表記の Q1-Q10)について、正答率の向上を調査した.

実験前と実験後のアンケートにおける 10 間(Q1-10)の平均正答率は、それぞれ 0.68(SD=0.11)、0.94(SD=0.04)であった。ここで、実験段階(実験前、実験後)とシステムモード(OneTime, Repeat)を因子とする二元配置分散分析を実施した。その結果、実験前後では有意な正答率の向上が見られた(F(1,32)=89.38、p<.001、generalized η^2 =.50)が、モードの違いでは有意な結果が得られなかった(F(1,32)=1.12、p=.30、generalized Φ =.50)のでは1.12、 Φ =.30、generalized Φ =.50)のでは1.12、 Φ =.30、generalized Φ =.50)のでは1.12、 Φ =.30、generalized Φ =.50)ののでは1.12、 Φ =.30、generalized Φ =.50)のでは1.12、 Φ =.30、generalized Φ =.50)のでは1.12 Φ =.30、generalized Φ =.50 Φ —.50 Φ =.50 Φ —.50 Φ =.50 Φ —.50 Φ —.50

alized η^2 =.02).

さらに、この各問について調査前後での正答率の変化を調査した.結果として、Q2を除く全ての設問で有意な正答率の向上が確認され、DualCheckの学習効果が高いことが確認できた.

次に、Q1–10 の類題(Q1a–Q10a)についても正答率を評価した。これらの設問は実験後にのみデータを収集したため、別途クラウドソーシングで収集した参照正答率 (N=50) を基準とした。

まず、この 50 名のクラウドソーシング参加者と元の調査の参加者の事前知識に大きな偏りがないか検証するため、一般正答率調査のサンプルとなった 50 人と、調査参加者の、調査開始前時点での Q1-10 の平均正答率を比較した。平均正答率は各々 0.73 (SD=0.16) と 0.68 (SD=0.11) であり、t 検定では、この 2 群間に有意な結果は得られなかった(t(49,33)=1.26, p=.21、Cohen's d=0.13)。従って、大きな偏りはないと判断し、Q1a=10a の比較を行った。

Q1a-10a の平均的な正答率は,個別データ収集調査と実験後アンケートでそれぞれ 0.72 (SD=0.14),0.94 (SD=0.04) であった.t 検定の結果,実験後アンケートの方が有意に正答率が高かった(t(49, 33)=8.49,p<.001, Cohen's d=0.70).二項検定の結果,10 問中6 間において有意差が確認され,Q2a を除く5 間でDualCheck 利用者の方が有意に正答率が高かった.以上から,ユーザ実験で使用した10 間(Q1-Q10)の正答率は,実験前よりも実験後の方が有意に高くなること,実験後のアンケートでのみ示される類題10間(Q1a-Q10a)の正答率は,DualCheck を利用していないユーザの正答率と比較して有意に高くなることが確認された.

4.2.2 ユーザビリティの評価

筆者らの先行研究では、DualCheck のユーザビリティについても評価した.まず、System Usability Scale(SUS) [21] で既存の CAPTCHA と DualCheck のスコアを一元配置分散分析で比較したところ、テキストベースの CAPTCHA や画像ベースの CAPTCHA に比べて、有意にスコアが高くなった.一方でチェックボックス形式の CAPTCHA よりはスコアが低くなったが、両者の有意差は確認されなかった.また、モードの違いによる SUS スコアの有意な差は確認されなかった.ここから、DualCheck のユーザビリティは、テキストベースの CAPTCHA や画像ベースの reCAPTCHA

のユーザビリティよりも有意に高くなることが確認された.

5. 知識定着検証のための追跡調査

5.1 調査の概要

前章までで述べた通り、DualCheck を用いて 2 週間 の学習をすることで、情報セキュリティ・倫理に関す る問題の多くの正答率が向上することが確認された. 一方で、DualCheck で出題された知識のがどの程度の 期間持続するかを確認することが、出題内容や同じ問題の出題頻度などを検討する上で重要である. そこで、前回のユーザ実験の調査から約 4 ヶ月後に、当時の参加者を対象としたテストを実施した. 以後この調査を追跡調査と呼ぶ.

5.2 調査の手順

まず,前回のユーザ実験参加者に対して追跡調査への参加を呼びかけた.前回参加者 34 名のうち 32 名(女性 25 名,男性 7 名,20 代 4 名,30 代 9 名,40 代 13 名,50 代 4 名,60 代 2 名)が参加した.なお,4.1節で示した前回のユーザ実験でOneTimeモードを利用した参加者は15 名,Repeatモードを利用した参加者は17 名であった.下記の調査を完了した参加者には、謝礼として税込330 円をクラウドソーシングサービスを介して支払った.

参加者にはまず、前回のユーザ実験において出題したものと同一の、情報セキュリティ・倫理に関する問題 30 問を解くように依頼した.この問題セットはQ1-10,Q1a-Q10aと、この 20 問のみに集中しないように加えた 10 問からなり、出題順はランダムとした.次に、DualCheck の利用後にインターネットリテラシーの変化や、身につけた知識が役立った事例があったか等を尋ねるウェブアンケートへの回答を依頼した.

6. 追跡調査の結果

6.1 参加者の正答率の推移と比較

筆者らの先行調査 [2] では、DualCheck 使用前と 2 週間の使用後それぞれでテストを出題し、正答率を測定した。これらの結果と今回の追跡調査の正答率を比較する。まず、15 日間ユーザ実験前の正答率と 4 ヶ月調査の正答率を比較する。Q1-Q10 の平均正答率を参加者ごとに集計したところ、ユーザ実験前、追跡調査でそれぞれ 0.69 (SD=0.17) と 0.91 (SD=0.13) であり、t 検定では、追跡調査の方が有意に平均値が高かった(t(31)=-5.85, p=<0.001, Cohen's d=1.46). さらに、

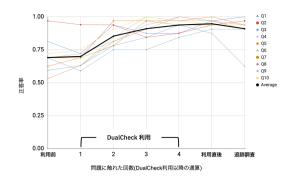


図 2 問題を解いた回数と、その正答率の推移を表すグラフ・横軸の 1-4 は DualCheck で問題を解いた回数を表す. 黒太線は、Q1-Q10 の平均正答率の推移を表す. / The accuracy transition across the number of exposure to the 10 questions. The black bold line represents the average accuracy.

問題ごとに正答率の二項検定を行った(表 1). このように、Q2 以外の 9 問において、追跡調査での正答率が 15 日間ユーザ実験前の正答率より有意に高かった. ここから、DualCheck 利用後 4 ヶ月経過後にも、DualCheck を利用して得られた知識が持続していると考えられる.

次に、15 日間ユーザ実験直後の正答率と追跡調査の正答率を比較する。Q1-Q10 の平均正答率はそれぞれ 0.95 (SD=0.08) と 0.91 (SD=0.13) であり、追跡調査の方が正答率が低いものの、t 検定では有意な差は確認されなかった(t(31)=1.98, p=.06, Cohen's d=0.11). また、Q1a-Q10a の平均正答率はそれぞれ 0.88 (SD=0.14) と 0.87 (SD=0.14) であり、追跡調査の方が正答率が低いものの、t 検定では有意な差は確認されなかった(t(31)=0.48, p=.63, Cohen's d=0.14). さらに問題ごとに正答率の二項検定を行った(表 2). 表 2 に示すように、Q5,9 及び Q5a,10a において追跡調査での正答率がユーザ実験直後の正答率より有意に低かった。なお、Q5,5a,10a については、二項検定の母比率が 1 であるために有意差が出ているが、いずれも追跡調査時に 94% 以上の高い正答率となっている.

図2は、Q1-Q10のそれぞれを解いた機会と、その際の平均正答率を表したグラフである. 横軸最右列が今回の追跡調査の正答率である. 問題ごとの正答率の平均(黒太線)を見れば、ユーザ実験最終日の正答率よりやや低下したものの、DualCheck 利用前と比較して高い正答率を維持していることがわかる.

	15 日間ユー ザ実験前の 正答率	追跡調査の 正答率	95% CI	p	
Q1	0.81	1.00	[0.89,1.00]	<.01	**
Q2	0.97	0.97	[0.84,1.00]	1.00	
Q3	0.69	0.91	[0.75,0.98]	<.01	**
Q4	0.69	0.94	[0.79,0.99]	<.001	***
Q5	0.63	0.94	[0.79,0.99]	<.0001	***
Q6	0.59	0.94	[0.79,0.99]	<.0001	***
Q7	0.69	0.94	[0.79,0.99]	<.0001	***
Q8	0.53	0.91	[0.75,0.98]	<.0001	***
Q9	0.59	0.63	[0.44,0.79]	0.86	
Q10	0.72	0.94	[0.79,0.99]	<.01	**

表 1 15 日間ユーザ実験前と追跡調査時の Q1-Q10 の正答率,及び両者の二項検定の結果. / The accuracies of Q1-10 observed in the pre- experimental questionnaire and the retention study. It also includes the binomial test result for each question.

6.2 一般正答率との比較

筆者らの先行研究 [2] において, DualCheck を利 用していないクラウドソーシング利用者を対象に. Q1-Q10,Q1a-Q10a の正答率を調査した. この正答率 を一般正答率と称する. Q1-Q10 の平均正答率を参 加者ごとに集計したところ,一般正答率,追跡調査 でそれぞれ 0.73 (SD=0.16) と 0.91 (SD=0.13) であ り, t 検定では, 追跡調査の方が有意に平均値が高かっ た (t(49,31)=-5.55, p=<.0001, Cohen's d=1.26). また, Q1a-Q10a の平均正答率はそれぞれ 0.72 (SD=0.14) と 0.87 (SD=0.14) であり、t 検定では、追跡調査の方が 有意に平均値が高かった. (t(49,31)=-4.80, p<.0001,Cohen's d=1.09). また, Q1-Q10, Q1a-Q10a について, 正答率の二項検定を行った(表3). 結果として全ての 問題で今回の追加調査の正答率が一般正答率を上回っ た. また、O1.3.4.5.6 の 5 問及び O4a.6a,7a,8a の 4 問 では有意な差が確認された.

6.3 アンケートへの回答

本追跡調査では、参加者に対して、DualCheckで身につけた情報セキュリティ・倫理に関する知識を実践的に用いているかを調査するアンケートを行った。アンケートは記名式で行われたが、以下では参加者のID(P1-P34)を用いて記す。

まず、DualCheck の利用後に、情報セキュリティ・倫理に関する意識が変わったかを選択肢形式で質問した。選択肢は「大きく変わった/少し変わった/全く変わらなかった」の3つであり、それぞれ1,28,3名が該当する選択肢を選んだ。従って、32名中29名(91%)が何らかの意識変革があったと回答したことになる。続

	15 日間ユー ザ実験直後 の正答率	追跡調査の 正答率	95% CI	p	
Q1	0.97	1.00	[0.89,1.00]	.63	
Q2	0.94	0.97	[0.84,1.00]	.72	
Q3	0.91	0.91	[0.75,0.98]	1.00	
Q4	0.97	0.94	[0.79,0.99]	.26	
Q5	1.00	0.94	[0.79,0.99]	< .001	***
Q6	0.97	0.94	[0.79,0.99]	.26	
Q7	0.94	0.94	[0.79,0.99]	1.00	
Q8	0.97	0.91	[0.75,0.98]	.08	
Q9	0.88	0.63	[0.44,0.79]	< .001	***
Q10	0.94	0.94	[0.79,0.99]	1.00	
Q1a	0.84	0.88	[0.71,0.96]	.81	
Q2a	0.88	0.94	[0.79,0.99]	.42	
Q3a	0.78	0.66	[0.47,0.81]	.09	
Q4a	0.91	0.88	[0.71,0.96]	.54	
Q5a	1.00	0.94	[0.79,0.99]	< .001	***
Q6a	0.78	0.72	[0.53,0.86]	.39	
Q7a	0.91	1.00	[0.89,1.00]	.07	
Q8a	0.97	0.91	[0.75,0.98]	.08	
Q9a	0.72	0.81	[0.64,0.93]	.33	
Q10a	1.00	0.97	[0.84,1.00]	< .001	***

表 2 15 日間ユーザ実験直後と追跡調査時の Q1-Q10, Q1a-Q10a の正答率,及び両者の二項検定の結果. / The accuracies of Q1-10, Q1a-Q10a observed in the post-experimental questionnaire and the retention study. It also includes the binomial test result for each question.

いて、DualCheck で身につけた知識が具体的に役立った場面があるかを選択肢形式で尋ねた. 回答は「あった/なかった」の2択で、それぞれ18名、14名が選択肢を選んだ.

続いて、上記の質問に関連して、具体的に意識が変わった場面を自由記述形式で回答してもらった。まず、 具体的に意識が変わった事例として、「メールの発信元の真偽を意識するようになった」[P17]、「災害時等に SNS で流れてくる現地の情報は、政府等の公式のものを除き疑って見るようになった。」[P33] などが挙げられていた。

次に,具体的に役立った事例については,「怪しいサイトにすぐ気付くようになった。」[P7] や「配送業などを語った迷惑メールを見極めることが出来た。SNSでなりすましアカウントを見つけ対処することが出来た。」[P30] のように,実際に迷惑メールや SNS のなりすましアカウント,偽サイトなどを見分けることができたと回答した参加者が複数名いた.

最後に、DualCheck 利用後に新たに学びたいと考えたことや十分理解できていない内容について尋ねた.「個人的な文書作成のものがどのような形で漏れてしまうのかを詳しく知りたい。」[P29] など、DualCheck

	一般正答率	追跡調査の 正答率	95% CI	p	
Q1	0.86	1.00	[0.89,1.00]	.02	*
Q2	0.86	0.97	[0.84,1.00]	.12	
Q3	0.72	0.91	[0.75,0.98]	.02	*
Q4	0.66	0.94	[0.79,0.99]	<.0001	***
Q5	0.68	0.94	[0.79,0.99]	<.001	***
Q6	0.48	0.94	[0.79,0.99]	<.0001	***
Q7	0.80	0.94	[0.79,0.99]	.07	
Q8	0.76	0.91	[0.75,0.98]	.06	
Q9	0.62	0.63	[0.44,0.79]	1.00	
Q10	0.82	0.94	[0.79,0.99]	.10	
Q1a	0.74	0.88	[0.71,0.96]	.11	
Q2a	0.98	0.94	[0.79,0.99]	.13	
Q3a	0.52	0.66	[0.47,0.81]	.16	
Q4a	0.64	0.88	[0.71,0.96]	<.01	**
Q5a	0.90	0.94	[0.79,0.99]	.77	
Q6a	0.20	0.72	[0.53,0.86]	<.0001	***
Q7a	0.80	1.00	[0.89,1.00]	<.01	**
Q8a	0.70	0.91	[0.75,0.98]	.01	**
Q9a	0.72	0.81	[0.64,0.93]	.33	
Q10a	0.98	0.97	[0.84,1.00]	.48	

表 3 Q1-Q10, Q1a-Q10a の一般正答率と追跡調査の正答率,及び両者の二項検定の結果. / The accuracies of Q1-10, Q1a-Q10a observed in the separate data collection study and the retention study. It also includes the binomial test result for each question.

で学んだことをより詳細に知りたいという意見が見られた他,「パソコンやスマホのウイルスについて、対策はどの程度必要なのか。SNS などで自分の紹介ページの写真にネット上の写真を保存して使うのは問題ないのか。」[P7] など,規範的な内容にとどまらず,より実践的な事項について学びたいと考える参加者が複数いることがわかった.

7. 考察

7.1 追跡調査の結果の分析

追跡調査の結果, 15 日間ユーザ実験直後に比べれば, クイズの正答率は一部有意に低下していたものの, 15 日間ユーザ実験前の正答率と比較して, Q2 と Q9 を除いて有意に高い正答率を維持していた. 従って, 4 ヶ月後の知識の持続効果は概ね確認できた.

Q2 は 15 日間ユーザ実験前より正答率が高く, 差を確認することが困難であったと考えられる. Q9 は以下に示す問題で, 正答は「B のみ正しい」である.

- A: 受信したメールが本物の銀行などから送られたものか確かめるためには、送信元アドレスの@の後ろを見ればよい。
- B: 送られてきた URL が本物か確認するために 重要なことの一つに、ドメインを確認することがある。

しかし、本追跡調査で誤答となった参加者の全員が「両方正しい」と回答していた. A は@の前後両方を確認したとしてもなお不十分であるため誤文であるが、この内容に気づかない参加者が A を正文と捉えたことが推察される. この結果から、参加者が A の文の内容を忘れていたことが考えられ、さらに問題文が誤答を誘発しやすいものである可能性も考えなければならない.

本研究では、出題した多くの問題において4ヶ月前の筆者らの先行研究[2]と同程度の正答率が確認された、本研究の結果により、先行研究[2]で確認された結果が偶然である可能性が低下するとともに、同研究で確認された知識定着効果が、短期的なものにとどまらず持続していることが明らかになった。従って、本研究によって、先行研究[2]のユーザスタディの設計や、そこから得られた知識定着効果に関する結論の妥当性を裏付けることができたと言える。

また、アンケート回答からの定性的なデータからは、 実際的な場面において DualCheck により提供された知識が役立ったケースが報告されていた. また、知識の習得のみならず、情報セキュリティ・倫理に対する関心を高めることにも DualCheck が貢献し得たことが伺えた. 今後の同様な学習システムの研究においては、情報セキュリティ・倫理に対する関心の高さがどの程度維持されたかの検証も重要であると考えられる.

7.2 本研究の制約

本研究には複数の制約が存在する。まず、調査の参加者の性別・年代の属性に偏りがあることが挙げられる。さらに、クラウドソーシングサービスで参加者を募集したことから、日常的にインターネットを使い慣れており、調査参加者のインターネットリテラシーが高かった可能性も想定される。今後は、例えばインターネットをあまり使わない人も対象とするなど、広範な参加者を募る大規模な調査を実施することが考えられる。また、今回扱った問題は全て、ユーザ実験内で少なくとも一度出題したことのある問題であった。今後は、一度も出題したことのない問題で、Q1-Q10と類似のテーマの問題を出題することで、問題の文言の変化等にも対応可能な知識が定着しているか測定することも有効だと考えられる。

8. 結 論

本研究では、ボット判定の機会を活用して情報セキュリティ・倫理の知識を提供する Dual Check につい

て、問題提示から 4 ヶ月後の知識定着度を測定した. 結果として、4 ヶ月が経過した後にも 15 日間のユーザ 実験前より高い正答率を示した. また、15 日間ユーザ 実験後と比較しても、正答率の平均は本題、類題ともに有意な低下はみられなかった. 本結果から、先行研究 [2] が示した知識定着効果が妥当なものであると示された. また、アンケートの結果からは、DualCheckで身につけた知識が実践的に役立つ事例があることも確認された. 今後は、より長期間が経過した際の正答率の推移を調査すること、多様な問題を出題して知識定着の程度の差を検証することなどが考えられる.

謝辞 本研究は、株式会社メルカリとインクルーシ ブ工学連携研究機構の共同研究である価値交換工学の 成果の一部です。また、調査の参加者の皆様と、貴重 な指摘を下さった矢谷研のメンバーに深謝いたします。

文 献

- [1] 独立行政法人情報処理推進機構, "「2019 年度情報セキュリティの倫理に対する意識調査」報告書," https://www.ipa.go.jp/files/000080783.pdf, 2019.
- [2] R. Yoshikawa, H. Ochiai, and K. Yatani, "DualCheck: Exploiting human verification tasks for opportunistic online safety microlearning," Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), pp.19–37, USENIX Association, Aug. 2022.
- [3] L. vonAhn, M. Blum, N.J. Hopper, and J. Langford, "Captcha: Using hard ai problems for security," Advances in Cryptology — EUROCRYPT 2003, ed. by E. Biham, pp.294–311, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [4] B. Reinheimer, L. Aldag, P. Mayer, M. Mossano, R. Duezguen, B. Lofthouse, T. vonLandesberger, and M. Volkamer, "An investigation of phishing awareness and education over time: When and how to best remind users," Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020), pp.259–284, USENIX Association, Aug. 2020.
- [5] 独立行政法人情報処理推進機構,"「2019 年度情報セキュリティの脅威に対する意識調査」報告書," https://www.ipa.go.jp/files/000080784.pdf, 2019.
- [6] G.A. Grimes, M.G. Hough, E. Mazur, and M.L. Signorella, "Older adults' knowledge of internet hazards," Educational Gerontology, vol.36, pp.173–192, 2010.
- [7] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L.F. Cranor, J. Hong, and E. Nunge, "Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish," Proceedings of the 3rd symposium on Usable privacy and security (SOUPS 2007), pp.88–99, 2007.
- [8] S. Sheng, M. Holbrook, P. Kumaraguru, L.F. Cranor, and J. Downs, "Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions," Proceedings of the SIGCHI conference on human factors in computing systems, pp.373–382, 2010.
- [9] U. Meyer and V. Drury, "Certified phishing: taking a look at public key certificates of phishing websites," Fifteenth Symposium on

- Usable Privacy and Security (SOUPS 2019), pp.211-223, 2019.
- [10] A. Trusty and K.N. Truong, "Augmenting the web for second language vocabulary learning," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, p.3179–3188, CHI '11, Association for Computing Machinery, New York, NY, USA, 2011. https://doi.org/10.1145/1978942.1979414
- [11] C.J. Cai, P.J. Guo, J.R. Glass, and R.C. Miller, "Wait-learning: Leveraging wait time for second language education," Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, p.3701–3710, CHI '15, Association for Computing Machinery, New York, NY, USA, 2015. https://doi.org/10.1145/2702123.2702267
- [12] T. Dingler, D. Weber, M. Pielot, J. Cooper, C.-C. Chang, and N. Henze, "Language learning on-the-go: Opportune moments and design of mobile microlearning sessions," ••, pp.••-••, 2017.
- [13] G.S. Mohammed, K. Wakil, and S.S. Nawroly, "The effectiveness of microlearning to improve students' learning ability," International Journal of Educational Research Review, vol.3, no.3, pp.32–38, 2018.
- [14] L. vonAhn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "recaptcha: Human-based character recognition via web security measures," Science, vol.321, no.5895, pp.1465–1468, 2008.
- [15] J. Yan and A.S. El Ahmad, "Usability of captchas or usability issues in captcha design," Proceedings of the 4th Symposium on Usable Privacy and Security, p.44–52, SOUPS '08, Association for Computing Machinery, New York, NY, USA, 2008. https://doi.org/10.1145/1408664.1408671
- [16] T. Yamamoto, T. Suzuki, and M. Nishigaki, "A proposal of fourpanel cartoon captcha," 2011 IEEE International Conference on Advanced Information Networking and Applications, pp.159–166, 2011.
- [17] V. Fanelle, S. Karimi, A. Shah, B. Subramanian, and S. Das, "Blind and human: Exploring more usable audio CAPTCHA designs," Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020), pp.111–125, USENIX Association, Aug. 2020.
- [18] N. Tanthavech and A. Nimkoompai, "Captcha: Impact of website security on user experience," Proceedings of the 2019 4th International Conference on Intelligent Information Technology, p.37–41, ICIIT '19, Association for Computing Machinery, New York, NY, USA, 2019. https://doi.org/10.1145/3321454.3321459
- [19] 文部科学省, "高等学校情報科「情報 i」教員研修用教 材第 1 章,"https://www.mext.go.jp/content/20200722mxt_jogai02-100013300_003.pdf, 2019.
- [20] 独立行政法人情報処理推進機構セキュリティセンター, " 情報セキュリティ 10 大脅威 2021," https://www.ipa.go. jp/files/000088835.pdf, 2021.
- [21] J. Brooke, "Sus-a quick and dirty usability scale," Usability evaluation in industry, vol.189, no.194, pp.4–7, 1996.

(xxxx 年 xx 月 xx 日受付)

吉川諒

2022 年,東京大学工学部電気電子工学科卒業.同年に東京大学大学院学際情報学府修士課程進学,在籍中.HCI・ユーザブルセキュリティに関する研究に従事.

落合秀也

昭和 58 年生. 平成 18 年東京大学・工・電子情報工学科卒. 平成 20 年同大学大学院・情報理工学系研究科・修士課程了. 平成 23 年同大学大学院・同研究科・博士課程了. 同年同大学 助教, 平成 29 年同大学大学院・情報理工学系研究科・准教授, 現

在に至る. 博士 (情報理工学,東京大学). IoT システム・プロトコル, P2P ネットワーク,遅延耐性ネットワーク,サイバー・セキュリティ,分散 AI 研究の他,IEEE1888 標準化活動,ISO/IEC JTC1/SC6 標準化活動に従事.東大グリーン ICT プロジェクト・ステアリング委員長. LAN セキュリティ監視プロジェクト,分散 AI プロジェクトを主導.

矢谷浩司

東京大学大学院工学系研究科電気系工学専攻准教授. 博士 (コンピュータ科学). ヒューマン・コンピュータ・インタラクション分野の幅広い研究に従事. セキュリティインタフェースの研究業績に対して,マイクロソフト情報学研究賞 (2022 年) 受賞.

Abstract Learning online safety and ethics is becoming critical for general Internet users. Our previous work demonstrated DualCheck. an opportunistic learning system about online safety and ethics that can be integrated into human-verification tasks in our previous research. Our 15-day user study confirmed the positive learning effect of DualCheck and its higher usability. In this paper, we conducted a retention study four months after the previous study. The participants in our retention study and showed significantly higher accuracy than before they participated in our previous study. The retention study also confirmed positive lasting effects of DualCheck.

Key words Information privacy and ethics, opportunistic learning, human-verification task