

著作物利用に関する注意

本著作物の著作権は特定非営利活動法人ヒューマンインタフェース学会に帰属します。本著作物は著作権者であるヒューマンインタフェース学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」に従うことをお願い致します。

引用情報 (<https://junkato.jp/ja/science>)

加藤 淳 (2018). ヒューマンインタフェース研究における再現性向上に向けた取り組み. ヒューマンインタフェース学会誌, 20 (1), 23-28.

研究再現性問題 特集

ヒューマンインタフェース研究における
再現性向上に向けた取り組み

加藤 淳 (かとう じゅん)

東京大学大学院 情報理工学系研究科博士課程修了。博士 (情報理工学)。2014 年より産業技術総合研究所研究員。Human-Computer Interaction 全般、とくにプログラミング体験 (PX) を向上させる研究に従事。ACM CHI 2013/2015 Honorable Mention など受賞。

詳細は URL に記載 : <https://junkato.jp/ja>

1. はじめに

本稿では、ヒューマンインタフェース (HI) 設計やヒューマンコンピュータインタラクション (HCI) のように、人を対象としたシステム構築と評価を行う学問分野における再現性向上に向けた取り組みを紹介したい。HCI は心理学 / 社会学とのオーバーラップが大きいですが、その議論は他の寄稿に任せることとする。HI / HCI 分野のユニークな研究手法として、自分たちで研究開発したシステムを用いた評価実験を行うところが挙げられる。評価実験の目的は、システムが意図通り動作することの検証だけでなく、システムによって初めて可能になったことの分析や、システムを利用することで人の行動がどう変容したかの調査など多岐に渡る。これは、実験の設計そのものや、実験から発見される現象がしばしば研

究上の主たる貢献となる実験心理学との大きな違いである。つまり、本稿で扱う HI / HCI 研究では単に再現性といっても「評価実験の再現性」と「システムの再現性」の2種類があって、これらは相当独立したものと考えられている。

2. 評価実験の再現性について

まず本章では、実験心理学などの分野と共通する「評価実験の再現性」について論じる。HI / HCI 分野の論文に評価実験を記載する場合は、どのような実験参加者をどのような方法で集めたか、どのような手順で実験を行ったかなど、結果に影響を与えそうな情報をなるべく詳しく記述する必要がある。これは実験心理学でも同様だろう。異なるのは、HI / HCI では自ら設計したシステムを自ら評価する点である。

そこで、我田引水にならないよう、実験の設計時や結果の考察時にバイアスが入らないよう気を付けなければならない。

2.1 定量評価

HI / HCI 分野における定量的な評価実験は、主に3つの目的で行われる。1つめは新技術の性能測定である。性能を客観的指標で測れる場合、工学的に精確な記述さえしてあればよく、再現性を担保するのは比較的容易である。2つめは人間の振る舞いを記述する数理モデルの妥当性検証である。たとえば Fitts' law^[1] という、マウスカーソルでターゲットをポインティングするのにかかる時間 T を見積もるための簡単な数理モデルがある。 T は、カーソル初期位置とターゲットの間の距離 D およびターゲットの大きさ W を用い

て $T = a + b \log_2 \frac{D}{W}$ と表される (a, b はデバイスに応じて決まる定数)。これも、実験参加者にポインティングのタスクを繰り返してもらえば評価できる。3つめは新しいインタラクション手法の使い勝手や印象を調べることであり、System Usability Scale (SUS^[2]) などのように確立された評価指標がある。定量評価では、これらの規定手順を踏むことで再現性を向上させ、バイアスを避ける努力がなされる。

定量的な評価実験に基づいて示された内容は、後続研究の論文で間接的に再現され、確かめられることが多い。既存のインタラクション手法を改善する研究は、既存手法を追実装したうえで、使い勝手を提案手法と比較する評価実験を行うことになる。Fitts' law はマウスカーソルが直線的にターゲットに到達するまでの時間を扱っているが、折れ線などジェスチャを描くのにかかる時間まで扱えるようにモデルを拡張した研究^[3]があり、この評価実験はオリジナル手法の再現実験を含んでいる。数理モデルを参考に設計されたシステムの評価実験にも、モデルの検証実験的な面がある。Bubble Cursor^[4] は、ターゲットの大きさ W をマウス位置に応じて動的に拡大するインタラクション手法である。マウスを近づけるほどターゲットを大きくすることで、ポインティングまでの移動量を小さく抑えることができる。この際、ポインティングに要する時間 T が、 W の代わりに最終的なターゲットの大きさ EW (Effective Width) を用いてかなり精確に見積れることが確認されており、これは提案手法の優秀さを示すとともに、Fitts' law をサポートする結果にもなっていると考えられる。

2.2 定性評価

HI / HCI 分野では、数値に現れない提案手法の得失を議論するために、定量評価だけでなく定性評価を併用する、あるいは定性評価のみ行うことも多い。詳細な方法論はインタラクションデザインに関する網羅的文献^[5]に譲るが、例えば、インタビューの結果を書き起こしたり、アンケートに設けた自由入力欄の記入内容を分析したりする。論文には、実験参加者の生の声を全部

載せることもないわけではないが、後続研究が参考にできるような抜粋と要約のみが載るのが通例である。これをもとに、提案手法の利点を確認するだけでなく、適用範囲の輪郭を明確にし、手法の限界と今後の課題まで明らかにすることを旨とする。また、著者の専門とする創作支援の研究では、実験参加者に提案システムを用いてコンテンツを作ってもらったワークショップ形式の実験を行うことも多い (具体例: [6, 7])。論文では、ワークショップの成果として得られたコンテンツを分類し、必要に応じて抜粋して紹介する。ワークショップ形式の実験では、実験参加者の発想次第で毎回異なるコンテンツが得られることになる。これを以て、再現性のない一回性の高い実験と批判することも可能だろう。

ここで、再現性が求められる「実験結果」の指すものについて考えてみたい。まず定量評価の場合、たとえ同じ手順を踏んだとしても誤差に阻まれて全く同じ統計量が得られるわけではない。そのため、有意差検定などの客観的評価を経て展開される定性的な結論こそが「実験結果」であろう。つまり、全く同じデータが得られるかどうか (例えば、提案手法を用いた際のタスク完了時間) ではなく、得られたデータに基づく議論 (例えば、提案手法がベースライン手法よりもタスク完了時間を短縮できたか) が再現できるかどうか (例えば、提案手法で作りやすいと主張されているコンテンツの種類が妥当かどうか) である。つまり、研究再現性問題における「実験結果」とは、論文の主張を支える具体的説得材料であり、研究者間で合意が取れてさえいれば、特定のフォーマットに縛られるものではない。

ただし論文読者は、評価実験のよしあしを再現性の有無だけで判断することはない。読者は初めから再現実験をしたいと思って論文を読むのではなく、むしろ実験結果を信頼して納得した上

で後続研究をしたいと考えるのが自然である。その際の信頼性向上の一つの方策として実験の再現性向上があるに過ぎず、他にも、実験を綿密に計画、実施していないと得られない面白い知見を報告したり、後述するようにシステムの再現性を向上したりといった多様な方策を組み合わせることができる。いかに論理の漏れのない説得的な組み合わせで実験を報告できるかが研究者の力の見せどころと言えよう。

2.3 どのような評価実験が必要か?

一般論として、定量評価にせよ定性評価にせよ、HI / HCI 研究の評価実験から本当に面白い知見を見出すことは容易ではない。評価実験を論文の「守備力」を上げるために必要な手続きとして軽視し、よく内容を吟味せず杓子定規に行っているケースが増えてきており、こうしたアライバイ作りのような実験については「むしろ害悪である^[8]」と批判されている。とくに新しいアルゴリズムやインタラクションデザインなどのアイデアを提案する研究では、アイデアそのものに研究上の価値が認められ、評価実験は不要と判断されることも珍しくない。こうした研究では、過去の研究からのインクリメンタルな差分を示すことよりも、新しく面白い問題を発見してエレガントに解くことのほうが重要視される。

HCI 分野のトップ国際会議は The ACM CHI Conference on Human Factors in Computing Systems (CHI) と The ACM Symposium on User Interface Software and Technology (UIST) が双璧をなす存在として認識されているが、前者が評価実験を重視する傾向にあるのに対し、後者は新規性やエレガントさを重視する傾向にある。こうした差は、前者が心理学者や社会学者も広く包含するコミュニティであるのに対し、後者はコンピュータグラフィクスやプログラミング言語など工学的な研究をしている人々から生じたコミュニティであるという歴史的経緯によって生じているものと考えられる。

このように、評価実験の再現性という概念が馴染まない研究もたくさんあるのが HI / HCI 分野である。当該分

野は極めて学際性が高く、研究の種類を3つの類型^[9]や6つの類型^[10]に分類して議論する論文があるほどで、それぞれの類型別に評価手法も全く異なるとされている。各論は参考文献に譲るとして、評価実験の設計に必要な戦略は、ある研究者曰く「最も自信がないところを検証するように」することだという。システムのいいところは、元からそうなるよう作ってあるので、それを検証しても驚くような結果はあまり出てこない。なるべく多くの情報を引き出す実験をするには、自分で自分の痛いところを進んで突くような、真摯で勇敢な姿勢が重要なかもしれない。

それでは、評価実験の適切な規模、具体的には実験参加者数についてはどう考えればよいのだろうか。例えば、CHIでよくある査読結果に「参加者数が少なすぎて評価実験が怪しい、やり直し」というものがある。確かに、参加者を増やしていけば実験結果のボリュームが大きくなり、参加者の隠れ変数的な偏りによる影響が小さくなるため、得られた結果を再現できる確証が増してくると考えられる。ただし、一方で査読者が言語化できない個人的な心理的抵抗感をこらした批判コメントに託していることがあり、素直に参加者数を増やせば問題が解消するとは限らない。そこで、適切な実験参加者数は実際に実験をしてみるまで分からないが、せめてコミュニティにおける業界標準を調べてみようとした調査研究がある^[11]。この調査研究によれば、CHI 2014採択論文では参加者数に1人から91万6千人まで開きがあり、12人の論文が最多であったという。結論としては、実験参加者が少人数でも論文として刊行できるが、その場合はとくに再現性が重要であること、どのような場合でもなぜその人数で十分と判断するに至ったか記述することなどを著者に推奨している。

2.4 評価実験の再現は可能か？

評価実験を再現する試みについては、それ単体では論文化しづらいこともあってなかなかトップ国際会議で目にするには少ない。数少ない例の一つが、クラウドソーシングを用いて先述したBubble Cursor^[4]を含む3種

類の実験を再現した取り組み^[12]である。いずれも元々の実験結果を再現する結果が得られており、クラウドワーカーの質や実験環境にばらつきがある場合でも、適切に実験用システムを設計すれば効率的に多くの被験者を集められることが分かっている。この研究はCHI 2013で発表され、RepliCHI Awardを受賞している。同じ著者のグループはLab in the Wild^[13]というクラウドソーシングプラットフォームを自ら立ち上げ、研究に活用している。

国際会議CHIにおいては、研究者有志によってHCI研究の再現性向上を目指すRepliCHI^[14]という取り組みが行われたことがある。CHI 2011でパネルセッション、CHI 2012でSpecial Interest Group (SIG) ミーティング、CHI 2013でワークショップ開催とRepliCHI Award設立まで行われたが、2014年でAwardが消失、2015年でワークショップも開催されなくなり、残念ながら定着しなかった。

ただ、評価実験についての議論はその後もしばしば話題に上っている^[15]。統計量は生データと比べると情報量が削減されているため、例えば全く同じ統計量を与える生データでも可視化してみると全く異なるグラフになることがある。この問題に着目し、元データ、統計量、グラフを入力として与えると、どう元データを操作すれば統計量を保ったまま目的のグラフが描けるか焼きなまし法で解いた研究がある^[16]。また、p値から有意差を議論することの危険性が議論されている^[17]。この研究グループが中心となって、論文にp値だけを載せるのではなく、より詳しい統計分布の情報まで載せることを推奨するTransparent Statistics in HCIという取り組みを推し進めている。こうすることで、著者らが統計の嘘をつけないだけでなく、後続研究が既存手法を再実装しなくても、以前の実験データをそのまま比較検討に利用できる利点があると主張している。

なお、心理学の分野では実験プロトコルを実験に先んじてジャーナルに投稿するPre-registrationという取り組みがあるという（本特集の三浦記事を参照）。しかしながら、HI/HCI分野ではプロトコルのなかに新規性のあるシ

ステムが含まれるため、これを忌避する研究者が多いのではないだろうか。理由としては、公知になっていないユーザインタフェースの詳細を投稿よりかなり前に査読者に渡してしまうと、アイデア盗用のリスクがあり、また、スケジュール上あまり現実的でないことが多いことが挙げられる。最近では複数グループが同時に実験を行う試みなどもあるそうだが、同様の理由でHI/HCIでは難しいだろう。

HI/HCI分野で心理学分野と同様の取り組みを行えない原因は他にもある。まず大前提として、実験を実施するためには、次章で述べるようにシステムの再現が必要となる。その上で、技術の日進月歩が激しく、実験条件が再現できなくなってしまうことが多い。例えばWebサイトを対象にした実験は、データセットを全部取っておくわけにいかないで再現できない。また、ハードウェアを対象にした実験は、パーツが入手できなくなったり、新型が出てサイズが合わなくなったりして再現できないことがある。さらに、システムの核となるアルゴリズムは再現できても、ソフトウェアやハードウェアの細部に渡るエンジニアリングの巧拙がユーザ体験に大きな影響を与えることがある。例えば、Microsoft OfficeやAdobe Creative Cloudといったプロダクトを拡張するような研究は、プロダクトの開発元にしかできないことがあり、社外の研究者による再現は不可能である。

3. システムの再現性について

HI/HCI研究の文脈で「再現性」というと、主たる対象は新しく提案されるユーザインタフェースやインタラクション手法、システムそのものである。本章では、その再現性を向上させるための取り組みについて紹介する。

3.1 ツールキット

今日使われているPC/Macのグラフィカルユーザインタフェース(GUI)は、1980年代に提案され、定着したものである。定着するまでは、個々の研究者がボタンやスライダーといったウィジェットを独自に実装していた。ある意味、再現実験をたくさんしてい

たのである。そのうち GUI ツールキットと呼ばれるウィジェットの実装をたくさん集めたライブラリが登場し、プログラマは new JButton() などと書くだけでボタンを画面上に表示できるようになった。ボタンをクリックするときに凹ませたり色を変えたりするような処理も、ツールキットが受け持ってくれる。

最近では、ソフトウェアだけでなくハードウェアのツールキット研究も多く提案されている。ソフトウェアであれば無償に近い低コストで配布できるが、ハードウェアの場合は製造コストがかかるため、研究を事業化するようなケースもある。こうした技術移転は、研究者にとってはかなり時間を取られる作業ではあるが、研究コミュニティだけでなく社会にも広く影響を与えるための手段として注目されている。例えば Phidgets は研究者以外でも購入でき、素人でもコンピュータを用いた電子工作 (Physical Computing) を楽しめる。類似のツールキットは数多く発売されており、市井の Maker 文化の醸成にも大きな役割を果たしている。

ツールキットは、既存手法を簡単に再現できるようにするだけでなく、その上で新しい研究を行うための道具の役割も果たしている。ツールキットの歴史は HCI の歴史であり、HI / HCI 研究者はツールキットを作り、使うことで、コミュニティ全体での車輪の再発明を避けながら社会にもインパクトを与えてきたのである^[18]。

3.2 オープンソース／オープンデータ

ツールキットほど標準化やパッケージ化を意識しなくとも、個々の研究プロジェクトで作ったシステムをオープンソースで公開するだけでも研究の再現性は向上できる。コンピュータ科学分野の最大学会である Association for Computing Machinery (ACM) は「Artifact Review and Badging」というタイトルで各分野の予稿集において「論文だけでなく対象となるシステムや取り扱ったデータ (artifact) を公開することを推奨し、一定の基準をクリアした論文にバッジを付与して讃えよう」というポリシーを表明している^[19]。Artifact Evaluation と呼ばれるこうした試みは ACM のなかでもソフトウェ

アに関する国際会議で始まり、現在はソフトウェアとデータセットが研究成果として認められるコンピュータシステム系の会議 (PLDI, OOPSLA, SC など) やコンピュータグラフィクスに関する会議 (SIGGRAPH) で行われている。この取り組みをサポートする会議では、プログラム委員会の他に Artifact Evaluation Committee という委員会が編成され、採録が決まった論文の中から再現性に関する一定の基準を満たしたものを選出してバッジを付与することになっている。

これらの Artifact Evaluation では、論文投稿者は GitHub などの誰でもアクセスできるリポジトリにすべてのソースコードとデータを置くことが一般的となっている。こうした取り組みは、近年の機械学習系の研究でも頻繁に見られる。とくに、機械学習などの分野では標準的なデータセットに対して分類精度などを競う研究が多いため、Kaggle^[20] のようにデータセットだけを掲載する Web サービスも出てきている。

オープンソース化の流れは、もちろん再現性向上に資することがモチベーションとなりうる。しかし研究者目線では、他の誰よりも早く研究成果を実現したことの証拠となるメリットの方が大きい。HI / HCI に限らずコンピュータ科学の諸分野では、ジャーナルの査読期間がかかりすぎるために査読付き会議が重宝されるようになった経緯があるが、それでも年に数回しかないため、流れの速い分野ではまずオープン化して後から査読プロセスを通そうという慣例もできつつあるようだ。

3.3 追実装／勉強会

たとえツールキット化されていなくても、オープンソースで公開されていなくても、論文の提案手法であれば、しっかり精読すれば再実装できるはずである。これは車輪の再発明を避けたがるコンピュータ科学分野では最後の手段ではあるが、極めて重要な取り組みである。そこで、例えば卒業論文や修士論文を書く前の準備段階で学生に明示的に論文の再現をさせる教授もいる。また、「CG 技術の実装と数理^[21]」のように、研究者同士で集まって追実

装をしたり実装について議論したりする会を開いていることもある。こうした取り組みが直接研究になることはないかもしれないが、論文に報告されていない詳細な事実 (例えば手法の思わぬ弱点など) が分かることは多く、後続研究に繋がることはとても多い。

さらに、そもそも論文が知られていなければ、論文に再現性が欠けていても誰も気づかないという問題がある。つまり、論文をより多くの人が知り、精読することが、再現性向上に間接的につながるのである。例えば一年に刊行される論文数が 600 本近い CHI や CVPR の全論文を紹介する「CHI 勉強会^[22]」や「cvcpaper.challenge^[23]」、諸分野に分散するプログラミング体験に関する論文を収集して情報交換する「SIGPX^[24]」といった勉強会が開かれている。

3.4 メタ科学としての HI / HCI

ここまで見てきたように、HI / HCI 分野における再現性向上のための取り組みは多層的で、こうすればよい、というはっきりした道筋が見えにくいことが多い。その根幹には「そもそも HI / HCI は一体どんな学問なのか?」という問いが横たわっている。これに答えるために、コミュニティ内ではメタ科学的な取り組みが行われてきた。例えば、HCI 研究を類型化して研究上の貢献の種類に応じて評価手法を別々に論じる研究^[7, 8]や、中でもツールキット研究に限定して進め方や評価手法について議論するワークショップ^[25]、既存のツールキット研究を詳しく分析したサーベイ^[26]などが行われてきている。こうした自己分析をメタ科学的なレイヤーで行うことは、学問分野の重要性を広く科学者コミュニティに伝えるための論点整理として非常に重要である。

一方で、我々 HI / HCI 研究者は人と人工物 / コンピュータのより良い関係を築く専門家である。自己分析から一歩進んで、コンピュータありきの現代における科学のあり方を建設的に提案することもできるはずである。コンピュータ科学の古典的エッセイ論文に「The Computer Scientist as Toolsmith^[27]」というものがある。HI / HCI 分野におけるツールキット研

究を持ち出すまでもなく、我々は人類の知的生産のための道具を作る応用科学者であり続けてきたのだ。例えばクラウドソーシングによって既存の実験を再現する Lab in the Wild^[13] は実験心理学者にも使ってもらえるだろう。PDF の論文に画像だけでなく動画を貼りこむ取り組みを自主的に進めていた研究者は、その内容を論文としてまとめ直し、より伝わる論文フォーマットについての議論に昇華させた^[28]。また、コンテンツ制作を Web ブラウザ上で完結させることで、制作に関わった全ての人を正確にクレジットでき、二次創作などが容易なコンテンツ制作環境^[29] や、研究用にも使える電子機器のソフトウェアとハードウェア、筐体設計まで一括開発できる Physical Computing 向け開発環境^[7] なども提案されており、こうした考え方や技術は科学研究に活用できるだろう。

ただ、研究の再現性を向上し、研究を加速するツールについて正面から取り組んだ研究は HI / HCI 分野ではまだ多くない。研究ノートとして開発された Jupyter Notebook^[30] は論文を執筆する目的でも使えるようになりつつあり、同様の取り組みは学術出版社 Elsevier が Executable Papers^[31] という名の下に進めている。研究用ツール開発に取り組むスタートアップ企業も増えてきている^[32]。今後、こうした社会の要請に応える HI / HCI 分野発のメタ科学研究が増えていくことを願っている。

4. おわりに

本稿では、HI / HCI 分野における再現性向上の取り組みについて紹介した。これが主たる縦串のトピックだとすると、横串は「HCI 分野の研究と実験心理学は何が違うのだろうか? HCI とは一体どんな学問なのか?」というメタ科学的なトピックである。折に触れて HCI 分野の研究手法を議論した文献を引き、また、HCI 分野が科学に対して貢献できる可能性についても議論した。読者諸氏におかれては、「The Computer Scientist as Toolsmith^[25]」の気概を持って、再現性の高い、信頼できる研究に取り組んでいただければ幸甚である。

謝辞

本稿は著者にとってメタ科学的な原稿をフォーマルに執筆する初めての機会であり、心理学、HI、HCI、Human-Robot Interaction、広くは自然科学の多分野に渡る研究者のみなさまからのフィードバックなくしては完成しなかった。ここに記して謝意を表する。

参考文献

- [1] Paul M. Fitts: The information capacity of the human motor system in controlling the amplitude of movement, *Jour. Exp. Psych.*, Vol.47, No.6, pp.381-391, 1954.
- [2] John Brooke: SUS: a 'quick and dirty' usability scale, *Usability Evaluation in Industry*, Taylor and Francis, London, UK, pp.189-194, 1996.
- [3] Johnny Accot, Shumin Zhai: Beyond Fitts' law: models for trajectory-based HCI tasks, In Proc. of CHI '97, ACM, New York, NY, USA, pp.295-302, 1997.
- [4] Tovi Grossman, Ravin Balakrishnan: The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area, In Proc. of CHI '05, ACM, New York, NY, USA, pp.281-290, 2005.
- [5] Jenny Preece, Helen Sharp, Yvonne Rogers: *Interaction Design: Beyond Human-Computer Interaction*, 4th Edition. John Wiley & Sons, Hoboken, NJ, USA, 2015.
- [6] Jun Kato, Daisuke Sakamoto, Takeo Igarashi: Phytobots: a toolkit for making robotic things, In Proc. of DIS '12, ACM, New York, NY, USA, pp.248-257, 2012.
- [7] Jun Kato, Masataka Goto: f3.js: A parametric design tool for physical computing devices for both interaction designers and end-users, In Proc. of DIS '17, ACM, New York, NY, USA, pp.1099-1110, 2017.
- [8] Saul Greenberg, Bill Buxton: Usability evaluation considered harmful (some of the time), In Proc. of CHI '08, ACM, New York, NY, USA, pp.111-120, 2008.
- [9] Antti Oulasvirta, Kasper Hornbæk: HCI research as problem-solving, In Proc. of CHI '16, ACM, New York, NY, USA, pp.4956-4967, 2016.
- [10] Jacob O. Wobbrock, Julie A. Kientz: Research contributions in human-computer interaction, *interactions*, Vol.23, No.3, pp.38-44, 2016.
- [11] Kelly Caine: Local standards for sample size at CHI, In Proc. of CHI '16, ACM, New York, NY, USA, pp.981-992, 2016.
- [12] Steven Komarov, Katharina Reinecke, Krzysztof Z. Gajos: Crowdsourcing performance evaluations of user interfaces, In Proc. of CHI '13, ACM, New York, NY, USA, pp.207-216, 2013.
- [13] Lab in the Wild: <http://www.labinthewild.org>
- [14] RepliCHI - revisiting, replicating, and reproducing HCI findings: <http://www.replichi.com>
- [15] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, Jakob Grue Simonsen: Is once enough?: on the extent and content of replications in human-computer interaction, In Proc. of CHI '14, ACM, New York, NY, USA, pp.3523-3532, 2014.
- [16] Justin Matejka, George Fitzmaurice: Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing, In Proc. of CHI '17, ACM, New York, NY, USA, pp.1290-1294, 2017.
- [17] Matthew Kay, Gregory L. Nelson, Eric B. Hekler: Researcher-centered design of statistics: why bayesian statistics better fit the culture and incentives of HCI, In Proc. of CHI '16, ACM, New York, NY, USA, pp.4521-4532, 2016.
- [18] Brad Myers, Scott E. Hudson, Randy Pausch: Past, present, and future of user interface software tools, *ACM TOCHI*, Vol.7, No.1, pp.3-28, 2000.
- [19] Artifact Review and Badging: <https://www.acm.org/publications/policies/artifact-review-badging>
- [20] Kaggle: <https://www.kaggle.com>
- [21] CG 技術の実装と数理: <http://cg.eng.shizuoka.ac.jp/~micgt2017>
- [22] CHI 勉強会: <http://sigchi.jp/seminar/chi2017>
- [23] Hirokatsu Kataoka, Yudai Miyashita, Tomoaki Yamabe, Soma Shirakabe, Shin'ichi Sato, Hironori Hoshino, Ryo Kato, Kaori Abe, Takaaki Imanari, Naomichi Kobayashi, Shinichiro Morita, Akio Nakamura: cvpaper.challenge in 2015 - A review of CVPR2015 and DeepSurvey, arXiv pre-print 1605.08247, 2016.
- [24] 加藤淳, 増原英彦 (編): プログラミング・エクスペリエンスの新潮流 - 言語設計から産業応用まで -, 情報処理, Vol.58, No.11 (2017年11月), pp.1006-1025, 2017.
- [25] Nicolai Marquardt, Steven Houben, Michel Beaudouin-Lafon, Andrew D.

- Wilson: HCITools: strategies and best practices for designing, evaluating and sharing technical HCI toolkits, In Proc. of CHI EA '17. ACM, New York, NY, USA, pp.624-627, 2017.
- [26] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, Saul Greenberg: Evaluation strategies for HCI toolkit research, <http://hdl.handle.net/1880/52215>, 2017.
- [27] Frederick P. Brooks, Jr.: The computer scientist as toolsmith II. Commun, ACM, Vol.39, No.3 (March 1996), pp.61-68, 1996.
- [28] Tovi Grossman, Fanny Chevalier, Rubaiat Habib Kazi: Bringing research articles to life with animated figures, Interactions, Vol.23, No.4 (June 2016), pp.52-57, 2016.
- [29] Jun Kato, Tomoyasu Nakano, Masataka Goto: TextAlive: integrated design environment for kinetic typography, In Proc. of CHI '15, ACM, New York, NY, USA, pp.3403-3412, 2015.
- [30] Fernando Perez, Brian E. Granger: IPython: A system for interactive scientific computing, Computing in Science & Engineering, Vol.9, No.3, pp.21-29, 2007.
- [31] Elsevier: Executable Papers - Improving the Article Format in Computer Science, <https://www.elsevier.com/physical-sciences/computer-science/executable-papers-improving-thearticle-format-in-computer-science>
- [32] 馬田隆明: 研究を加速するスタートアップ, <https://www.slideshare.net/takaumada/startup-tool-for-research>