Memory and Motor Processes of Password Entry Error

Franklin P. Tamborello, II (franklin.tamborello.ctr@nrl.navy.mil)
National Reseaerch Council Postdoctoral Research Associate
Washington, DC, USA

Kristen K. Greene (kristen.greene@nist.gov)
National Institute of Standards and Technology[1]
Gaithersburg, MD, USA

Passwords are tightly interwoven with the digital fabric of our current society. Unfortunately, passwords that provide better security generally tend to be more complex, both in length and composition. Complex passwords are problematic both cognitively and motorically, leading to both memory and motor errors during recall and entry. It is important that we better understand and disentangle the two error sources, as password entry errors can have significant negative consequences, such as being locked out of a critical information system. We present a computational cognitive model of password recall and typing, with memory and motor errors each contributing to password entry error. With this synthesis we can study human-computer interaction issues involving the usability of computer access control systems, specifically the password as an authentication mechanism. Ultimately we hope to make science-based recommendations for password policies that promote the use of passwords that are more usable.

INTRODUCTION

Despite widespread recognition that character-based passwords are a deeply problematic method of user authentication (Honan, 2012), they are tightly interwoven with the digital fabric of our current society. The ubiquity of passwords is true both for personal and work place accounts, as is the challenge of complying with a variety of password policies (Shelton, 2014; Choong & Theofanos, 2015). People are forced to remember—or in some other way keep track of—a large and ever-increasing number of passwords as they interact with a variety of systems and accounts each day (Florencio & Herley, 2007; Choong, Theofanos, & Liu, 2014).

In addition to an increasing number of passwords, people must also contend with passwords of increasing length. Computer security specialists suggest increasing the length of passwords; this increases their entropy, or randomness, which makes them more computationally expensive to guess. Furthermore, passwords are increasing in complexity as well as length. For most systems—particularly systems in higher-security enterprise environments—passwords containing only lowercase letters are not permitted. In addition to lowercase letters, the inclusion of uppercase letters, numbers, and special characters is also required, as using all four character categories is often recommended for increasing password security (United States Department of Homeland Security, 2009).

Most password requirements also prohibit the use of words, as dictionary attacks on passwords are so successful, even since the late 1970s (Morris & Thompson, 1979). This means that higher-entropy passwords can differ greatly from the natural language words used in studies on skilled typing and transcription typing (e.g., Coover, 1923; Gentner, 1981; Salthouse, 1984; Salthouse, 1986). While words follow orthographic rules and are predictable given neighboring semantic content, passwords should ideally be as random as possible to help mitigate guessing. While non-word strings of random letters have been included in prior transcription typing research (e.g., Salthouse, 1984), the numbers and special characters suggested for passwords were not.

Although there are longterm research efforts underway to replace passwords (National Strategy for Trusted Identities in Cyberspace, 2011), widespread implementation will take some time. Furthermore, even as newer identity management systems and authentication technologies such as biometrics become more prevalent, legacy systems may remain reliant upon passwords. Therefore, balance between usability and security in password policies remains important.

Unfortunately, due to privacy and security concerns, it can be difficult to collect real-world password data. To collect laboratory data from large numbers of participants across a variety of password requirement combinations would require prohibitively large investments of time and money. Usable security is certainly not the only domain where access to human data can be challenging, and as in other domains, computational cognitive modeling offers a promising alternative to augment existing behavioral research.

Drawing upon theories from cognitive science and experimental psychology can help understand the roles that human cognition and motor movement play in generating, rehearsing, recalling, and typing passwords on various devices. Unifying theories of memory and motor error can help inform recommendations for password policies that better address both the limits and capabilities of human performance. By supplementing behavioral data from prior password studies with predictive models of human performance, we can test theories and hypotheses in ways that neither research method can do alone.

In particular, we are interested in whether existing theories and models can disentangle memory from motor errors for those complex, system-generated passwords suggested or required in higher-security enterprise

[1] Disclaimer: Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products mentioned are necessarily the best available for the purpose.

environments. In such enterprise environments, passwords differ from words in several important ways, which means that traditional memory, transcription typing, and mobile text entry literature and theory may not be completely sufficient to inform and test predictive models of password typing. The usable security literature may address this somewhat, yet many password studies do not report sufficiently detailed data for model validation purposes.

REVIEW

There have certainly been many studies on memory in general (e.g., Miller, 1956; Baddeley & Hitch, 1974; Unsworth & Engle, 2007), and password memorability in particular (Vu, Bhargav-Spantzel, & Proctor, 2003; Forget & Biddle, 2008; Chiasson et al., 2009). There is also a large existing body of literature examining expert typing and transcription typing from the 1920s to the 1980s (e.g., Coover, 1923; Gentner, 1981; Salthouse, 1984; Salthouse, 1986). There has been a comprehensive cognitive model of transcription typing, Bonnie John's TYPIST model (1996), which quantified 19 of the 29 phenomena reviewed by Salthouse (1986), as well as two additional phenomena. However, these studies did not include stimuli similar enough of complex passwords to suit our modeling goals.

Although the typing literature and models do well at examining the cognitive and perceptual-motor facets of typing, there are certain distinctions between passwords and words that may not be fully addressed by existing theory and research. For example, the cost of errors and error recovery can differ significantly between typing for communicative purposes, such as composing emails, and typing for authentication tasks (i.e., password entry). Typos in communication can be embarrassing, but typos in passwords can cause failed authentication attempts, which in turn cause accounts to be locked. Users are sensitive to the time (and frustration) cost of unlocking an account, which may impact their speed-accuracy tradeoff function specifically for password entry in comparison with other text entry tasks. This may be particularly true on mobile devices, where users cannot rely on the now common predictive algorithms for password entry. There is a rich body of mobile text entry literature examining factors such as the effect of devices (Castellucci & MacKenzie, 2011), motion (Nicolau & Jorge, 2012), and age (Nicolau & Jorge, 2012) on how people type words or phrases, but again, such stimuli are not representative of the complex passwords we are interested in modeling.

One important difference between general text entry and password entry is the lack of visual feedback during password entry tasks. On desktop computers, text is masked immediately as it is typed. On mobile devices, the character just typed is generally visible for a moment[2] before being masked. An additional difference between general text entry and complex password entry is the required navigation back and forth between multiple onscreen keyboards that password entry requires of the user. Passwords requiring a number of onscreen keyboard changes, or screen depth changes, can have disproportionately large effects across both entry times and error rates (Greene, Gallagher, Stanton, & Lee, 2014).

Studies using password-like stimuli and masked text can help to address the aforementioned literature gaps and provide much-needed data to inform computational cognitive models of the often onerous password entry task. There have been both desktop (Stanton & Greene, 2014) and mobile studies (Greene, Gallagher, Stanton, & Lee, 2014; Gallagher, 2015) using such complex password-like stimuli. As our current focus is on modeling desktop password entry errors, we focus much of our review on the desktop study and model that motivated our work.

Stanton and Greene (2014) examined the usability of system-generated passwords by having participants memorize a series of ten passwords and type them repeatedly using a desktop computer. Participants were given one password at a time. For each password, there was a set of three task phases: practice, verification, and entry. During the practice phase, participants could practice typing the password as many (or as few) times as they wished. The password was visible, and typed text was also visible during the practice phase. During verification, typed text was still visible, but the password was not. Participants had to enter the memorized password correctly during the verification phase in order to move on to the entry phase. During the entry phase, participants had to type the the memorized password ten times. After the series of three phases (practice, verification, and entry) was completed for each of the ten passwords, there was a surprise recall test. For the surprise recall test, typed text was visible.

The Stanton and Greene (2014) study examined the fundamentals of desktop password typing, contributing baseline data on human performance with stimuli representative of the complex, system-generated passwords found in higher-security enterprise environments. Most relevant for the current work were Stanton and Greene's (2014) error findings: at 45% of the total error corpus, incorrect capitalization errors were by far the most prevalent. Incorrect capitalization, or shifting, errors were almost three times as likely as the next most prevalent error category (missing character errors, or omissions, were 17% of the total error corpus).

The nature of the most common error category (incorrect capitalization, or shifting errors) is interesting for several reasons. The high frequency of incorrect capitalization errors was particularly important given the fact that most modern password policies—and certainly those in higher-security enterprise environments—require at least one uppercase letter. Additionally, most special characters (which are also required by many password policies) require a shift action. Twenty-one of the total 32 possible special characters require shifting; only 11 special characters can be executed without requiring a shift action. Finally, of greatest interest for our modeling efforts is the fact that based purely on the behavioral data reported in Stanton and Greene (2014), it cannot be fully determined whether those errors were memory errors or motor execution errors (or a combination of both).

Greene and Tamborello (2015) began modeling work to disambiguate memory from motor errors using a single password from the Stanton and Greene (2014) stimuli set. They report a cognition-only ACT-R model of password

[2] This is a setting that can be changed; for increased security to help protect against shoulder-surfing, mobile keyboard settings allow the user to turn the momentary visibility feature off for password text fields.

rehearsal, finding that recall errors alone were insufficient to fully explain the incorrect capitalization errors of interest. They also report an expansion of ACT-R's native typing abilities to support password-specific typing needs, giving ACT-R the ability to type capital letters and symbols, and to err while doing so. Such modifications were necessary to explore the role of motor error during desktop password entry, as the canonical ACT-R architecture is limited to perfect typing performance and would not predict the motor execution errors expected with typing complex passwords. Furthermore, the canonical ACT-R architecture does not support case-sensitivity in typing, nor does its typing vocabulary support all possible symbols; without such capacity, it would be impossible to model typing complex passwords.

ACT-R

We use the ACT-R cognitive architecture (Anderson et al, 2004) to model user password recall and typing. ACT-R is a hybrid symbolic and subsymbolic computational cognitive architecture that takes as inputs knowledge (both procedural and declarative about how to do the task of interest) and a simulated environment in which to run. It posits several modules, each of which perform some aspect of cognition (e.g., long-term declarative memory, vision). Each module has a buffer into which it can place a symbolic representation that is made available to the other modules. ACT-R contains a variety of computational mechanisms and the output of the model is a time stamped series of behaviors including individual attention shifts, speech output, button presses, and the like. It can operate stochastically and so models may be non-deterministic.

NEW CONTRIBUTION

Our model works by incorporating and coordinating two distinct systems underlying prospective memory and motor operations. The former operates on the principle of associative spreading activation (Anderson et al., 2004) while the latter builds upon the motor models embodied in EPIC (Meyer & Kieras, 1997A & B) and ACT-R (Anderson et al, 2004).

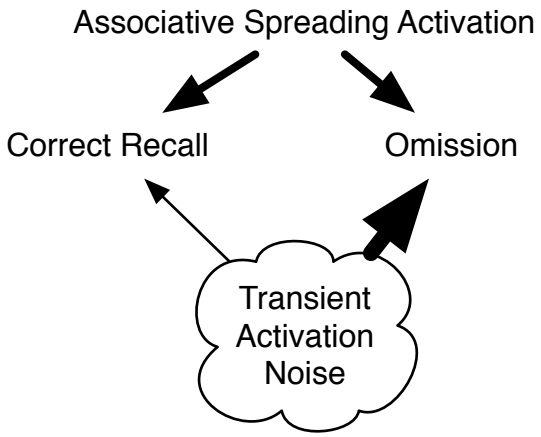

Figure 1. The role of noise in the model's memory processes: Associative spreading activation is the prospective memory process underlying selection of correct actions. When transient activation noise, a fundamental property of human memory, spikes during prospective retrieval it can lead to an omission.

Password Sequence Recall

Sequential tasks require prospective memory to remember what comes next. Our model uses this memory process, selecting the next character using the current character to prime retrieval.

Selecting the next character. Sequence memory is a prospective memory task, using a representation of the current character to associatively prime retrieval of a memory representation of the next character. We use ACT-R's spreading activation mechanism to implement prospective memory. Furthermore, activation propagates from active buffer contents to long-term memory according to what we assume to be learned association from each context to its subsequent action (Botvinick & Plaut, 2004).

Memory Errors

Memory errors arise out of the interaction of noise with the processes of normal task execution (Figure 1).

Omission. We assume that association is somewhat imprecise in that there is not a clean one-to-one mapping of cue to target. Instead, some association "bleeds" over from the target to a handful of subsequent items, with each subsequent item receiving less association than the one coming before it in sequence. The model may omit a character when transient noise is such that it simultaneously suppresses activation of the correct next step and enhances activation of one of these subsequent items.

Investigating the source of password entry errors is a perfect application opportunity for cognitive modeling to shed light on the root cause of error that was intractable to ascertain through prior behavioral data alone. By implementing support for an ACT-R model that can type capital letters, one could then test different models to see whether those incorrect capitalization errors were memory errors or motor execution errors (where a shift key press had been attempted but simply not executed properly, such as by prematurely releasing the shift key). The ability to type capital letters raises interesting theoretical questions. For each letter of the alphabet, do people have two distinct versions in their memory, one lowercase and one uppercase? Or is an uppercase letter encoded as the lowercase plus a required shift action?

Implementation Issues in ACT-R

In order to support modeling of incorrect capitalization typing errors, two limitations in ACT-R first required addressing: missing special characters and lack of case-sensitivity in typing.

Missing Special Characters. Of the non-alphanumeric characters available on typical American English keyboards, ACT-R previously included support only for the period, semicolon, slash, and quote (Bothell, 2014, see "key" on page 320 of the ACT-R Reference Manual). Therefore, in order to enable modeling typing of the remaining special characters, we added support for all remaining ASCII printable characters not previously supported by ACT-R.

Lack of Case-Sensitivity. Regardless of whether calling ACT'R's "press-key" motor module request (Bothell, 2014, see page 317 of the ACT-R Reference Manual) with a capital or lowercase letter, the output will be the same in ACT-R's current instantiation. This is somewhat problematic for modeling incorrect capitalization errors, which requires that ACT-R be capable of press-and-hold capability for the left and right shift keys, combined with a simultaneous key press of a second key (i.e., chorded typing). Therefore we added to ACT-R a capability to type key chords and output case-sensitive text, as described in the following section.

Stochastic Typing Extension for ACT-R

The standard ACT-R distribution (Anderson, et al, 2004; Anderson 2007) does not predict any typing errors as a matter of motor error (Bothell, 2014). However, real humans, even very skilled typists, are imperfect, and tend to err at rates from 0.5% to 35% (Salthouse, 1986; Panko, 2008; Landauer, 1987). We wished to explain password entry errors, but because some errors are due to memory processes and some are due to motor processes, we had to extend our modeling framework of choice, ACT-R, so that it, too, would be capable of such motor errors. Furthermore, we needed to implement the low-frequency, non-alphanumeric characters that information systems often require their users to incorporate into their passwords as a matter of security policy, e.g. "*" or "?". Source code for the ACT-R stochastic typing extension may be downloaded from https://github.com/usnistgov/CogMod.

Motor Errors in Typing

Our typing extension for ACT-R redefines some of ACT-R's existing code so that any requested typing action can stochastically result in the output of a typed key other than the one intended. It adapts the ellipsoid motor movement error equation of May (2012) and Gallagher and Byrne (2013), producing greater error along the axis of movement than off the axis, the off-axis error being scaled to .75 of the on-axis. However, because here the units are keys rather than pixels as in May's study, and ACT-R assumes most keys are the same width, the width term in May's equation is simplified to 1.

Hold-Key. Because typing non-alphanumeric characters typically involves holding a shift key while striking another key, and standard ACT-R provides no way to hold any such modifier key, it was necessary to invent such a method. Our errorful typing extension provides two motor module request extensions (see "extend-manual-requests" on page 325 of the ACT-R Reference Manual, Bothell, 2014) to enable the holding and releasing of modifier keys such as shift.

The new hold-key motor module request acts like press-key, translating the requested key to be held into a peck movement (Bothell, 2014, pp. 315-316) with the appropriate features. Once the hold-key motor movement is executed, ACT-R will have a state indicating that the appropriate key is being held. This state in turn causes ACT-R to now output a different character for the same press-key requests that follow for the given keys. The model can request the release-key function to release the given modifier key and end the modifier key state.

Nonalphanumeric Characters. With a shift key held, ACT-R can now type non-alphanumeric ASCII characters such as "*" and "?." It can now also type capital letters as well as lower-case letters, a critical feature for case-sensitive passwords lacking in standard ACT-R.

DISCUSSION

As in other domains, computational cognitive modeling can be a useful tool in the usable security research field, where behavioral data from prior password studies can be supplemented with predictive models of human performance. Although the study that motivated our work was focused on passwords for higher-security enterprise environments, our work has implications beyond that restrictive environment. By extending a widely used cognitive architecture to address motor errors in a way it previously did not, we contribute to the growing corpus of typing models (e.g., John, 1988; John, 1996; Das & Stuerzlinger, 2007; Gallagher, 2015; Gallagher & Byrne, 2015; Greene & Tamborello, 2015), all of which act together to test and expand the ACT-R theory.

Memory Errors

The kinds and frequencies of sequence memory errors arise from the fundamental properties of that memory system. Work on this problem from other domains (e.g. Anderson et al, 2004; Botvinick and Plaut, 2004) lend strong support to the memory account we use here, associative spreading activation.

Motor Errors

Motor errors are their own important contributor to password entry error, as the shifting errors in Stanton and Greene's (2014) study so strikingly exemplify. Moreover, as mobile touchscreen computers continue to gain importance it will become necessary to understand the mechanics of motor errors involved with that interface and how they contribute to password entry errors.

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological Review, 111(4), 1036-60. doi:10.1037/0033-295X.111.4.1036

Anderson, J. R. (2007). *How can the human mind exist in the physical universe?* New York, NY: Oxford University Press. Retrieved from Google Scholar.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In Bower, G. (ed.) Recent Advances in Learning and Motivation, vol. 8, pp. 47–90. Academic Press, New York.

Bothell, D. (2014). ACT-R 6.0 reference manual. ACT-R Research Group. Retrieved from act-r.psy.cmu.edu

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. Psychol Rev, 111(2), 395-429. doi:10.1037/0033-295X.111.2.395

Castellucci, S. J., & MacKenzie, I. S. (2011). Gathering text entry metrics on android devices. In CHI 2011 Extended Abstracts on Human Factors in Computing Systems, pp. 1507–1512.

Coover, J. E. (1923). A method of teaching typewriting based upon a psychological analysis of expert typing. National Education Association 61, 561–567

Chiasson, S., Forget, A., Stobert, E., Van Oorschot, P., & Biddle, R. (2009). Multiple password interference in text

passwords and click-based graphical passwords. In Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 500–511.
Choong, Y., Theofanos, M., & Liu, H. (2014). United States Federal Employees' Password Management Behaviors - a Department of Commerce Case Study. National Institute of Standards and Technology Interagency Report (NISTIR) 7991.
Choong, Y., & Theofanos, M. F. (2015). What 4,500+ People Can Tell You – Employees' Attitudes toward Organizational Password Policy Do Matter. To Appear in Proceedings of the 3rd International Conference on Human Aspects of Information Security, Privacy and Trust, in the 17th International Conference on Human-Computer Interaction.
Das, A., & Stuerzlinger, W. (2007). A cognitive simulation model for novice text entry on cell phone keypads. Proceedings of the 14th European Conference on Cognitive Ergonomics: invent! explore!, 141-147.
Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. In Proceedings of the 16th International Conference on World Wide Web (WWW), pp. 657-666. ACM, New York.
Forget, A., & Biddle, R. (2008). Memorability of persuasive passwords. In CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 3759–3764.
Gallagher, M. A. (2015). Modeling Password Entry on Mobile Devices: Please Check Your Password and Try Again. Doctoral Dissertation, Rice University, Houston TX.
Gallagher, M. A., & Byrne, M. D. (2015). Modeling Password Entry on a Mobile Device. To appear in Proceedings of the 2015 International Conference on Cognitive Modeling.
Gallagher, M. A., & Byrne, M. D. (2013). The devil is in the distribution: Refining an ACT-R model of a continuous motor task. In *Proceedings of the 12th International Conference on Cognitive Modeling*. Ottawa, Canada.
Gentner, D. (1981). Skilled finger movements in typing. Center for Information Processing, University of California, San Diego. CHIP Report 104
Greene, K. K., Gallagher, M. A., Stanton, B. C., & Lee, P. Y. (2014). I Can't Type That! P@$$w0rd Entry on Mobile Devices. In Human Aspects of Information, Security, Privacy, and Trust. Lecture Notes in Computer Science, Vol. 8533, pp 160-171.
Greene, K. K., & Tamborello, F. P. (2015). Password Entry Errors: Memory or Motor? To appear in Proceedings of the 2015 International conference on Cognitive Modeling.
Honan, M. (2012). Kill the password: Why a string of characters can't protect us anymore. Wired.
John, B.E. (1988). Contributions to Engineering Models of Human-Computer Interaction, Department of Psychology, Carnegie-Mellon University, Ph.D. thesis.
John, B.E. (1996). TYPIST: A theory of performance in skilled typing. Human Computer Interaction, 11, 321-355.
Landauer, T. K. (1987). Relations between cognitive psychology and computer systems design. In J. M. Carroll (Ed.), Interfacing thought: Cognitive aspects of human-computer interaction (pp. 1-25). Cambridge, MA: MIT Press.
May, K. (2012). A model of error in 2D pointing tasks. Undergraduate Honors Thesis, Rice University, Houston, TX.
Meyer, D. M., & Kieras, D. K. (1997). A computational theory of executive control processes and human multiple-task performance: Part 1. Basic mechanisms. Psychological Review, 104, 3-65. Retrieved from Google Scholar.
Meyer, D. M., & Kieras, D. K. (1997). A computational theory of executive control processes and human multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. Psychological Review, 104, 749-791. Retrieved from Google Scholar.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63(2), 81–97. Morris, R., & Thompson, K. (1979). Password Security: A Case History. Communications of the ACM, 22(11): 594-597.
Morris, R., & Thompson, K. (1979). Password Security: A Case History. Communications of the ACM, 22(11): 594-597.
National Strategy for Trusted Identities in Cyberspace. Enhancing Online choice, Efficiency, Security, and Privacy. (2011). Retrieved online from http://www.whitehouse.gov/sites/default/files/rss_viewer/NSTICstrategy_041511.pdf
Nicolau, H., & Jorge, J. (2012). Elderly text-entry performance on touchscreens. In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, Boulder.
Nicolau, H., & Jorge, J. (2012). Touch typing using thumbs: understanding the effect of mobility and hand posture. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2683–2686.
Panko, R. R. (2008.). Basic error rates. [Web page] Retrieved from http://panko.shidler.hawaii.edu/HumanErr/Basic.htm
Salthouse, T. (1984). Effects of age and skill in typing. Journal of Experimental Psychology 113(3), 345–371
Salthouse, T. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. Psychological Bulletin 99(3), 303–319.
Shelton, D. C. (2014). Reasons for Non-Compliance with Mandatory Information Assurance Policies by a Trained Population. Doctoral Dissertation, Capitol Technology University.
Stanton, B. C., & Greene, K. K. (2014). Character Strings, Memory and Passwords: What a Recall Study Can Tell Us. In Human Aspects of Information, Security, Privacy, and Trust. Lecture Notes in Computer Science, Vol. 8533, pp 195-206.
United States Department of Homeland Security. (2009). United States Computer Emergency Readiness Team (US-CERT), Security tip (ST04-002): Choosing and protecting passwords. Retrieved online from http://www.us-cert.gov/cas/tips/ST04-002.html
Unsworth, N., & Engle, R. W. (2007). The foundations of remembering: Essays in honor of Henry L. Roedgier III, pp. 241-258. Psychology Press, New York.
Vu, K., Bhargav-Spantzel, A., & Proctor, R. (2003). Imposing password restrictions for multiple accounts: Impact on generation and recall of passwords. In Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society (HFES), pp. 1331–1335.