A New Behavioral Measure of Cognitive Flexibility

Christian A. Gonzalez[1], Ivonne J. Figueroa[1], Brooke G. Bellows[1],
Dustin Rhodes[2], and Robert J. Youmans[1]

[1] George Mason University, Department of Psychology, Fairfax, Virginia, USA
cgonza12@gmu.edu
[2] University of California, Santa Cruz, California, USA

Abstract. Individual differences in cognitive flexibility may underlie a variety of different user behaviors, but a lack of effective measurement tools has limited the predictive and descriptive potential of cognitive flexibility in human-computer interaction applications. This study presents a new computerized measure of cognitive flexibility, and then provides evidence for convergent validity. Our findings indicate moderate to strong correlations with the Trail Making Task, and in particular, those aspects of the task most closely associated with cognitive flexibility. Results of this study provide support for the validity of a new measure of cognitive flexibility. We conclude by discussing the measure's potential applicability in the field of HCI.

Keywords: cognitive flexibility, individual differences, user modeling.

1 Introduction

When a user interacts with a system, they bring along their unique skills, biases and abilities. In the past, models of individual differences have been used to predict user behavior [1, 2], however, these models often lack a cognitive component. Conversely, cognitive modeling has been successful in designing, planning and evaluating systems for expert users [3, 4], but typically does not reflect the impact of individual differences in cognitive abilities [5]. *Cognitive flexibility* (CF), defined as an a person's ability to abandon one cognitive strategy in favor of another based on a change in task demands [6], represents one individual difference that may underlie a variety of different user behaviors.

One way to understand the importance of CF is to examine the behaviors that are associated with its absence, namely, *perseveration*. Extreme perseveration is defined as a maladaptive repetition of a particular behavior, and is a well-studied phenomenon in clinical psychology and neuropsychology [7–9]. Outside of clinical populations, more mild perseverative tendencies also naturally occur, and these impact a wide range of everyday activities. Research suggests that CF predicts behaviors ranging from how likely a person is to notice changes in their environment [10] to how creative a poem they are likely to write [11]. Most importantly here, we believe that CF has potential applications towards improving human-computer interaction (HCI).

For example, a recent study [12] found that locating "hard-to-find" features was one of the major sources of frustration during a computer interaction task. Individual differences in CF may explain why some users are more likely to find the sorts of hidden features that others miss because CF predicts an ability to disengage from a specific search or quickly abandon inefficient strategies. Finally, research has identified CF as one of the primary mechanisms of insight when solving problems [13, 14]. Some speculate this is because "it benefits from 'cognitive restructuring' of the problem, enabling the solver to pursue a new strategy or a new set of associations [15]." Overall, how users differ in their approach to problems, and their ability to change cognitive strategies, is relevant to HCI; however, without an effective measure of CF, both its descriptive and predictive potential in HCI will remain limited.

Currently, CF is often measured using two well-established tasks, the Wisconsin Card Sorting Task (WCST) and the Trail Making Task (TMT) [16–18]. In the WCST, the participant's goal is to sort a series of cards according to one of three rules: shape, color, or number. Participants begin unaware of which rule is active, then must learn the sorting rule in response to experimenter feedback, and finally must reacquire a new sorting rule when the old one changes. The WCST variable, 'percent perseverative errors', is most often associated with CF [19]. This variable is a ratio of the number of number of errors attributed to perseveration over the total number of errors made. The TMT assesses flexibility slightly differently. First, a baseline score is obtained in part A where, participants make a 'trail' by connecting an ascending series of numbers. Then, in part B, an additional series of letters is added and participants are required to connect the ascending series alternating between the two. Scores on part B and the difference between part B and part A, are most often associated with CF [18]. In general, the WCST measures errors and the TMT measures increases in time both caused by a lack of CF. However, neither task is really ideal for measuring both: Time on trials when rule switches occur are typically not compared to non-switch trials in the WCST and errors on the TMT are only reflected in increased completion time caused by fixing the error [16, 18]. In addition, both tasks have established themselves as part of neuropsychological batteries used for diagnosing executive dysfunction, but have not seen as wide acceptance as measures of cognitive abilities among healthy populations. By comparison, measures of constructs like working memory capacity, such as the operation span task [20, 21] have been used extensively in individual difference research in a variety of different populations including HCI studies [22, 23].

More recently, self-report methods of CF have been developed. The cognitive flexibility scale (CFS) created by Martin and Rubin [24, 25] measures flexibility in the context of effective communication. However, the CFS approaches the concept of CF differently than behavioral measures by dividing the construct into three areas: awareness of alternatives, willingness to be flexible and self-efficacy in being flexible. The CFS was validated with other measures of communication effectiveness and found to be internally reliable with high test-retest reliability ($r = .83$). A more recent self-report measure, the cognitive flexibility inventory (CFI), builds upon the CFS and extends its utility [26]. The CFI applies to more general life situations and was

intended for clinical populations to support cognitive behavioral therapy (CBT) for patients with depression. The authors found that cognitive inflexibility, as measured by the CFI, was associated with more depressive symptoms measured by the Beck Depressive Inventory (BDI-II; r = -.39 time 1 and -.37 time 2). The CFS and CFI were also found to be highly correlated (r = .73 time 1 and r = .75 time 2). Overall, the self-report measures of CF suggest that there is a conscious aspect of flexibility related to recognizing alternatives and choosing to act on them. However, there is currently little research comparing results of self-report measures and behavioral measures of CF like the WCST and TMT.

2   A New Measure of Cognitive Flexibility

The goal of this research is to develop a measure of CF that draws from both the WCST and the TMT in order to reliably measure individual differences in flexible thinking in normal populations. By establishing a comprehensive measure of CF, we hope that fields like HCI will be able to assess and incorporate individual differences in CF into predictive and descriptive models of user behavior. Furthermore, in an effort to find consensus between behavioral and self-report measures of CF, we the TMT and the CFS with our measure. The measure presented here is a computerized version of a paper-and-pencil puzzle task developed by the second and last author [27]. In their study, Figueroa and Youmans found that the WCST variable, 'Trials to Complete First Category,' was a significant predictor of puzzle completion time such that fewer trials to complete first category (negatively related to CF) was associated with faster puzzle completion times.

However, as previous work has described [28], the paper puzzle was limited in many ways. The only dependent variable produced was a single puzzle completion time, allowing only for indirect inferences about the specific impact of rule switching on that variable. Unlike the paper-based puzzle, the computerized version presented here (Figure 1) allows for multiple puzzles, manipulation the number of switches per puzzle and measurement of switch and non-switch move times. By administering multiple puzzle trials we were able to assess how the amount of switching per trial impacted each individual. In addition, the possibility of 'dead-ends', a concern with the paper puzzle, was eliminated in the computerized version. Some additional aesthetic differences between paper and computerized versions include: different shapes and colors, a fog-of-war that occludes all tiles except current and previous moves and a nineteen-move path compared to the paper puzzle's twenty-two move path.

In this study, we attempt to further validate our measure of CF by correlating performance with the TMT as well the CFS. We hypothesized that performance on the puzzle task would indicate an individual's cognitive flexibility because participants must maintain an active rule to move quickly through the puzzle on non-switch moves, but must also quickly abandon previous rules in order to make progress on switch moves. As a consequence, perseveration on the previous rule should lead to increased switch move times. Furthermore, we reasoned that if switch costs, the time differences between switch and non-switch moves, were robust [29–31], then puzzle

completion times would increase with the number of switches in a given puzzle. Thus, a simple bivariate regression with number of switches predicting completion time would allow us summarize the average effect increasing switches had on each participant's performance. We expected that scores on the TMT B and derived scores would be positively correlated with puzzle performance in terms of both faster puzzle trial completion times and reduced switch cost, and that CFS scores would be negatively correlated with performance.
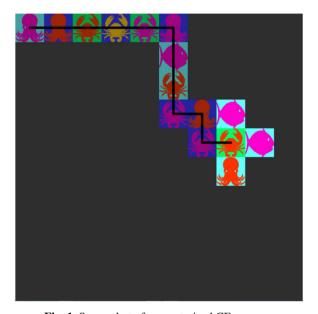

Fig. 1. Screenshot of computerized CF measure

3   Method

3.1   Participants

Twenty-four participants (7 men and 18 women, between 18 and 30 years old, median 19) from George Mason University's undergraduate research pool voluntarily participated for class credit.

3.2   Materials and Procedure

These data were collected as part of a larger study investigating how CF affects internet search. The TMT, CFS and Puzzle were all administered in this order immediately after participants had completed a series of Internet search tasks.

Trail Making Task. Paper-and-pencil-based versions using Reitan's (1955) arrangement of the TMT parts A and B were administered. In part A, participants connected

a series of circles numbered 1-25 in order. In part B, participants connected an alternating series of numbers and letters (e.g. 1 to A, 2 to B, 3 to C etc.). Scores on the TMT were in the form of completion times as measured by the experimenter using a stopwatch. Errors were accounted for in terms of time required to correct. Direct scores and derived scores (i.e. B-A and B/A) were used for analysis. Participants always completed part A before part B.

Cognitive Flexibility Scale. Participants completed the 12-item CFS. Items (e.g. "I am willing to work at creative solutions to problems") were scored on a 6-point scale of agreement ("strongly agree" to "strongly disagree"). Higher scores indicated greater levels of flexibility (maximum of 72).

Puzzle. The computerized puzzle was completed in Adobe® Flash®. Participants used a mouse to navigate a 10 x 10 grid of tiles (60x60 pixels each). Each tile had a specific shape, shape color and background color. A "fog-of-war" occluded all moves except those immediately available and participants' previously traveled path, limiting the amount of planning (Figure 1). Of immediately available moves (either two or three depending on the location in the puzzle) there was ever only one correct, legal move, eliminating the possibility of dead ends. In order to make a legal move, participants needed to match their current tile to the desired tile by three different rules: shape, shape color and background color. As participants completed the task, the active matching rule changed and the participant was required to adopt the new rule in order to continue. For example, a participant might make three successive moves by matching by background color, then on the fourth move, no tiles match the background color of the current tile, forcing the participant to abandon the background color matching rule and adopt a new rule based on the tiles available i.e. matching by either shape or shape color. Participants completed the nineteen-move path by always starting from the top-left corner and ending in the bottom-right corner. Each puzzle was randomly generated, with switch moves and order of presentation randomized to mitigate any order effects. Participants completed seven puzzle trials, each containing between two and fourteen switches.

Participants viewed a training PowerPoint and completed three practice puzzles with the experimenter to ensure they fully understood how to navigate the puzzle properly. Before continuing with the experiment, all participants were trained to the criterion that they were able to complete a two-switch, eight-move practice puzzle within thirty seconds. The task and training were run using a Macintosh iMac with a 21.5-inch screen.

The computerized puzzle allowed for the measurement of several different variables. We measured the completion times for each puzzle in seconds, average switch and non-switch move times, and the additive, linear effect of amount of switching on completion times (*b*). Each participant completed 56 switch and 77 non-switch moves across all seven puzzle trials. Switch moves occurred when the participant switched to a new rule in order to advance. Non-switch moves occurred when they advanced according to the rule in the previous move. Logically, any move that was not a switch move would be non-switch move and vice versa. Switch cost was derived by

calculating the difference between average switch and non-switch move times for each puzzle then averaged across trials for each participant. This variable indicates the additional time required to switch rules over and above just making a move. We expected all raw and derived scores to be positively correlated with TMT B, B-A and B/A and negatively correlated with scores on the CFS.

4   Results

After data were screened one outlier, two standard deviations above the mean, was removed. The following analysis was conducted on the remaining twenty-three participants. However, since a participant could have a small b value when fitting a line through a set of highly varied data points, we also accounted for the adjusted R^2 ($\bar{R}^2$) of the line of best fit. For this reason, only b values corresponding to an $\bar{R}^2$ value of .30 or higher were used for analysis. We acknowledge that fitting a linear model over seven data points violates regression assumptions necessary for obtaining linear unbiased estimates, so the reader is encouraged to interpret b values as summary statistics rather than inferential or predictive in any way. After applying the $\bar{R}^2$ cutoff, eight participants were excluded, leaving only fifteen participants analyzable b values. Correlations for this variable are specific to those fifteen participants, but all other variables will refer to the full sample of 23.

Five variables, TMT B, B-A, B/A, Switch Cost and Switch move time violated the assumption of normality (Shapiro-Wilk $p < .05$) required for calculating Pearson product-moment correlations. After performing natural log transforms, all five variables, except TMT B were approximately normal (Shapiro-Wilk $p > .05$). Pearson product-moment correlations were computed between TMT, CF, puzzle dependent variables (DVs) and are shown in Table 1.

Internal consistency of the puzzle trials was assed via Cronbach's alpha ($\alpha = .89$; bootstrap 95% CI [.80, .93]). Three of the five derived puzzle DVs had moderate to strong, significant positive correlations with TMT B and B-A scores providing evidence for convergent construct validity of the puzzle. In addition, five of the seven raw puzzle completion times were significantly correlated with TMT B and B-A. Switch cost was significantly correlated with TMT B scores but not B-A. Interestingly, switch move time was the only DV not significantly correlated with TMT or other puzzle DVs. Average puzzle completion time as well as the six and twelve switch puzzle completion times were significantly positively correlated with TMT A. B/A scores were not significantly correlated with any puzzle DVs. Though generally in the expected direction, correlations between the CFS and TMT and puzzle DVs did not reach significance.

5   Discussion

The goal of this study was to validate a new measure of cognitive flexibility. The high Cronbach's alpha suggests good internal consistency across trials and the strong positive correlations with the TMT provide evidence for convergent validity. Our results

Table 1. Correlations between TMT CFS and Puzzle

	TMT				CFS					Puzzle							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
TMT A																	
TMT B[1]	0.75**																
B - A[1]	0.34	0.87**															
B/A[1]	-0.25	0.43*	0.81**														
CFS	-0.03	-0.16	-0.23	-0.26													
b (n=15)	0.35	0.64**	0.63*	0.39	0.09												
Avg. Time	0.47*	0.71**	0.65**	0.34	-0.1	0.67**											
Switch Cost[1]	0.33	0.41*	0.37	0.19	-0.02	0.39	0.26										
Switch[1]	0.39	0.36	0.26	0.04	-0.33	0.27	0.32	-0.03									
Non-Switch	0.29	0.55**	0.54**	0.34	0.07	0.60*	0.81**	0.42*	-0.26								
2switch	0.21	0.39	0.40	0.26	-0.34	0.10	0.68**	0.01	0.21	0.51*							
4switch	0.33	0.35	0.28	0.04	-0.31	0.24	0.63**	0.00	0.37	0.38	0.45*						
6switch	0.53**	0.73**	0.63**	0.30	0.13	0.71**	0.84**	0.27	0.02	0.85**	0.36	0.47*					
8switch	0.24	0.54**	0.58**	0.41	-0.09	0.73**	0.81**	0.39	0.25	0.73**	0.40	0.34	0.66**				
10switch	0.40	0.61**	0.55**	0.29	-0.03	0.61*	0.88**	0.33	0.24	0.73**	0.57**	0.50*	0.75**	0.68**			
12switch	0.34	0.55**	0.52*	0.30	-0.05	0.51	0.79**	-0.05	0.44*	0.50*	0.73**	0.29	0.51*	0.66**	0.65**		
14switch	0.48*	0.66**	0.57**	0.24	0.04	0.88**	0.83**	0.41	0.20	0.72**	0.39	0.46*	0.77**	0.60**	0.72**	0.52*	

yielded three major findings. First, the puzzle correlated with those TMT dependent variables most closely associated with CF and executive function, TMT B and B-A, but did not correlate as strongly with TMT A, which has been found to reflect more basic motor and perceptual abilities [18]. Furthermore, the fact that the *b* values, were positively correlated with TMT B and B-A scores is a critical finding. This allows us to better disentangle flexibility from confounding aspects of CF like visual motor abilities (a common criticism of the TMT [18, 32, 33]).

Second, we did not find a correlation between the behavioral measures and the CFS. This suggests that perhaps these measures tap separate aspects of CF[1]. However, this study was too limited in scope to draw defensible conclusions about the overall relationship between behavioral and self-report measures of CF. As Dennis and Vander Wal [26] suggest, more research is needed in this area. A more comprehensive study with a larger sample and a wider range of measures would be better suited to answering this research question.

Third, this study attempted to characterize flexibility in terms of the linear effect of switching on completion time. However, approximately 30% of our sample had an unacceptably low linear fit ($\bar{R}^2$ below .3). This finding may suggest that a segment of the population does not experience a linear additive effect of switching. Perhaps there is a majority of individuals that demonstrate a good linear fit and a range of possible *b* values corresponding to high and low flexibility, and a separate group of individuals that may be distracted, de-focusing their attention or adapting to the task in an unforeseen way. This presents challenges for measurement, but allows for interesting speculation about what sets the 'poor-fit' individuals apart.

Better measures of individual differences like CF allow for the inclusion of cognitive abilities during interaction. Accounting for the flexibility of users with the measure presented here may allow for better prediction of which users are most susceptible to perseverative and ultimately potentially frustrating behaviors, which is a major a concern for designers [12]. Furthermore, cognitive models used in simulations may be able to use data from our measure to predict behaviors of average, high and low flexibility users, identifying what aspects of an interface or task require flexible thinking and how individual differences in flexible thinking may impact performance.

The study presented here highlights only the first step in an effort to bring the study of individual differences in cognitive ability to the field of HCI. The limitations of the human attentional system play a vital role in crafting technologies that are functional and easy to use [36]. Numerous studies have demonstrated that expertise, personality, age and gender may all impact user interactions [34, 35]. However, though the role of cognition in HCI is readily apparent, the role of individual differences in cognitive ability is not. Perhaps the most important principle of design, well-known to many in the field of HCI is 'know the user.' This simple aphorism presents an exceedingly difficult task. Understanding users' cognition and how that may vary across individuals will play a critical role in designing interfaces and experiences for a growing population of users in years to come. We hope that better tools and additional research will lead to a more complete understanding of user behavior and interaction with technology.

---

[1] Anecdotally, Dennis and Vander Wal have unpublished data documenting a similarly null relationship between behavioral and self-report measures of CF.

References

1. Harrison, A.W., Rainer Jr, R.K.: The influence of individual differences on skill in end-user computing. Journal of Management Information Systems, 93–111 (1992)
2. Hong, W., Thong, J.Y.L., Wong, W.M., Tam, K.Y.: Determinants of user acceptance of digital libraries: an empirical examination of individual differences and system characteristics. Journal of Management Information Systems 18, 97–124 (2002)
3. Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Commun. ACM. 23, 396–410 (1980)
4. Gray, W.D., John, B.E., Atwood, M.E.: Project Ernestine: Validating a GOMS Analysis for Predicting and Explaining Real-World Task Performance. Human-Computer Interaction 8, 237–309 (1993)
5. Olson, J.R., Olson, G.M.: The growth of cognitive modeling in human-computer interaction since GOMS. Human-Computer Interaction 5, 221–265 (1990)
6. Scott, W.A.: Cognitive complexity and cognitive flexibility. Sociometry, 405–414 (1962)
7. Allison, R.S.: Perseveration as a sign of diffuse and focal brain damage. I. Br. Med. J. 2, 1027–1032, contd (1966)
8. Müller, J., Dreisbach, G., Brocke, B., Lesch, K.P., Strobel, A., Goschke, T.: Dopamine and cognitive control: The influence of spontaneous eyeblink rate, DRD4 exon III polymorphism and gender on flexibility in set-shifting. Brain Research 1131, 155–162 (2007)
9. Eslinger, P.J., Grattan, L.M.: Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. Neuropsychologia 31, 17–28 (1993)
10. Youmans, R.J., Figueroa, I.J., Kramarova, O.: Reactive Task-Set Switching Ability, Not Working Memory Capacity, Predicts Change Blindness Sensitivity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 55, 914–918 (2011)
11. Figueroa, I.J., Youmans, R.J.: Individual differences in cognitive flexibility predict poetry originality. In: Proceedings of the 15th International Conference on Human-Computer Interaction, Las Vegas, Nevada (2013)
12. Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., Shneiderman, B.: Determining causes and severity of end-user frustration. International Journal of Human-Computer Interaction 17, 333–356 (2004)
13. Beversdorf, D.Q., Hughes, J.D., Steinberg, B.A., Lewis, L.D., Heilman, K.M.: Noradrenergic modulation of cognitive flexibility in problem solving. Neuroreport 10, 2763 (1999)
14. Baas, M., De Dreu, C.K., Nijstad, B.A.: A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? Psychological Bulletin 134, 779 (2008)
15. Subramaniam, K., Kounios, J., Parrish, T.B., Jung-Beeman, M.: A brain mechanism for facilitation of insight by positive affect. Journal of Cognitive Neuroscience 21, 415–432 (2009)
16. Reitan, R.M.: The relation of the Trail Making Test to organic brain damage. Journal of Consulting Psychology 19, 393–394 (1955)
17. Reitan, R.M.: Validity of the Trail Making Test as an indicator of organic brain damage. Perceptual and Motor Skills 8, 271–276 (1958)
18. Sanchez-Cubillo, I., Perianez, J.A., Adrover-Roig, D., Rodriguez-Sanchez, J.M., Rios-Lago, M., Tirapu, J., Barcelo, F.: Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. Journal of the International Neuropsychological Society 15, 438 (2009)
19. Barceló, F., Knight, R.T.: Both random and perseverative errors underlie WCST deficits in prefrontal patients. Neuropsychologia 40, 349–356 (2002)

20. Turner, M.L., Engle, R.W.: Is working memory capacity task dependent? Journal of Memory and Language. Journal of Memory and Language 28, 127–154 (1989)
21. Unsworth, N., Heitz, R.P., Schrock, J.C., Engle, R.W.: An automated version of the operation span task. Behavior Research Methods 37, 498–505 (2005)
22. Zander, T.O., Kothe, C., Jatzev, S., Gaertner, M.: Enhancing human-computer interaction with input from active and passive brain-computer interfaces. Brain-Computer Interfaces, 181–199 (2010)
23. Wong, A.W.K., Chan, C.C.H., Li-Tsang, C.W.P., Lam, C.S.: Competence of people with intellectual disabilities on using human–computer interface. Research in Developmental Disabilities 30, 107 (2009)
24. Martin, M.M., Rubin, R.B.: A new measure of cognitive flexibility. Psychological Reports 76, 623–626 (1995)
25. Martin, M.M., Anderson, C.M.: The cognitive flexibility scale: Three validity studies. Communication Reports 11, 1–9 (1998)
26. Dennis, J.P., Vander Wal, J.S.: The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. Cognitive Therapy and Research 34, |241–253 (2010)
27. Figueroa, I.J., Youmans, R.J.: Developing an Easy-to-Administer, Objective, and Valid Assessment of Cognitive Flexibility. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 944–948 (2011)
28. Gonzalez, C., Pratt, S.M., Benson, W., Figueroa, I.J., Rhodes, D., Youmans, R.J.: Creating a Computerized Assessment of Cognitive Flexibility with a User-Friendly Participant and Experimenter Interface. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1942–1946 (2012)
29. Monsell, S.: Task switching. Trends in Cognitive Sciences 7, 134–140 (2003)
30. Rogers, R.D., Monsell, S.: Costs of a predictible switch between simple cognitive tasks. Journal of Experimental Psychology: General 124, 207 (1995)
31. Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A.M., Koch, I.: Control and interference in task switching—A review. Psychological Bulletin 136, 849 (2010)
32. Crowe, S.F.: The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. J. Clin. Psychol. 54, 585–591 (1998)
33. Gaudino, E.A., Geisler, M.W., Squires, N.K.: Construct validity in the Trail Making Test: what makes Part B harder? Journal of Clinical and Experimental Neuropsychology 17, 529–535 (1995)
34. Helander, M.G., Landauer, T.K., Prabhu, P.V.: Handbook of human-computer interaction. North Holland (1997)
35. Aykin, N.M., Aykin, T.: Individual differences in human-computer interaction. Computers & Industrial Engineering 20, 373–379 (1991)
36. Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. CRC (1986)