

# Evaluating the Usability of System-Generated and User-Generated Passwords of Approximately Equal Security

Sourav Bhuyan, Joel S. Greenstein, and Kevin A. Juang

Clemson University

sbhuyan@g.clemson.edu, {iejsg,kjuang}@clemson.edu

**Abstract.** System-generated and user-generated text-based passwords are commonly used to authenticate access to electronic assets. Users typically have multiple web accounts ranging from banking to retail, each with a different password, creating a significant usability problem. The passwords authenticated by these applications may vary in usability and memorability depending on the type of password generation, composition and length. Researchers have compared the usability of different user-generated password composition schemes. The passwords created using different composition schemes in these studies achieved different levels of minimum security, making comparisons across them difficult. This research compares the usability and memorability of three password generation schemes that each exceed a specified minimum entropy for the sake of security.

**Keywords:** passwords, usability, security.

## 1 Introduction

The earliest passwords were generated by a computer system and assigned to the employees to ensure overall security [1] [2]. However, as they were composed of apparently random characters having no meaning for the users, they were more difficult to remember than user-generated passwords [3]. This high degree of complexity caused users to externalize them by writing them down, leading to potential breaches in security [3]. It led to user-generated passwords becoming widely used [1] even though system-generated ones are more difficult to guess [3]. To enhance the security of user-generated passwords, they can be created using a large domain of character sets, giving them the appearance of being randomly generated [3]. However, password guidelines that encourage users to do this, though they may help to create passwords that are more difficult to crack, also become difficult to use [4]. The limitations associated with restrictions on user-generated passwords include the time needed to generate an acceptable one, the guidelines that result in less memorable ones than those generated without them and the additional restrictions that may cause more entry errors and lengthen the login procedure [5]. This issue concerning password generation is made more complex because users also tend to form their own mental models of good passwords regardless of the instructions provided, favoring memorability

over security [6]. As a result, users circumvent password guidelines when given a chance, meaning that their passwords are still subject to being breached by brute force attacks. In such attacks, the intruder uses a computer to systematically attempt all possible combinations using a standard US keyboard of 94 characters [7]. In order to protect against such attacks, password guidelines recommend the use of all character sets and longer passwords [7].

The increased use of the Internet has led to an increase in the number of password applications [4]. Users now have multiple web accounts ranging from banking to retail, each with a different password [4], creating a significant usability problem [8]. To improve security and usability of user-generated passwords, proactive password checking is frequently implemented to ensure that user-generated passwords satisfy the composition guidelines [5]. These composition guidelines generally constrain user-generated passwords with respect to length, composition of character sets and inclusion in a dictionary [9].

More recently, researchers have compared the usability of different user-generated password composition schemes. However, the passwords created using different composition schemes in these studies achieved different levels of minimum security, making comparisons across them difficult. To expand on this research, this study compared passwords satisfying NIST Level 2 [10] security requirements that were either assigned by the system or created by the user using two different composition schemes.

## 2 Related Studies

To compare the usability and preferences of user-generated passwords and randomly assigned passwords, Zviran et al. [3] had 103 participants create two user-generated passwords in addition to being assigned an eight-character random password. One of the user-generated passwords was a maximum of 8 characters long and the other was an alphanumeric passphrase of up to 80 characters. After three months, the participants' recall success rate was the highest for the 8-character user-generated passwords, followed by assigned random passwords and finally the 80-character passwords. These results were supported by the data obtained with a subjective questionnaire in which the participants ranked the 8-character user-generated passwords highest in appeal and ease of recall. These passwords were further analyzed to determine the characteristics affecting their recall. The results revealed that 92% were composed of only lower case letters, suggesting better memorability of passwords of this composition.

To understand the effect of various composition schemes and additional guidelines or restrictions on password usability, Proctor et al. [5] conducted an experiment involving 24 participants. For the first condition, the participants created a password of at least 5 characters, and for the second, passwords incorporated the additional guidelines of having at least one member from all the character sets on a keyboard, at most

one character from the username and no consecutive repeated characters. The results indicated that the passwords with additional composition restrictions were significantly harder to generate and remember than those based on the minimal requirements. However, it was observed that the passwords with minimal requirements were weaker than the ones with additional requirements after all the passwords were subjected to password cracking software.

Using a similar procedure, Proctor et al. [5] conducted a second experiment which required a minimum length of at least 8 characters for the passwords. Similarly, the results of the second experiment found a statistically significant difference between the time taken to generate and recall minimal condition passwords and those requiring additional guidelines. Also similarly to the first experiment, the qualitative data found that passwords with additional guidelines were significantly harder to generate and remember compared to passwords with only a length restriction. The results concerning the breached passwords from both experiments suggested that the increase in the minimum length of minimal condition passwords from 5 to 8 characters led to passwords that were as resistant to password cracking software as the minimum 8-character password incorporating additional guidelines.

A more recent study investigating various user-generated password construction schemes was conducted by Vu et al. [11]. They investigated the number of attempts and the time required to generate passwords. They also evaluated the number of login errors and the time required to recall these passwords after a short and long duration of time. Results of the experiments conducted in the study indicated that user-generated passwords composed of initial letters of at least six words of a meaningful sentence were significantly easier to generate compared to similarly composed passwords that additionally included a number and a special character. The results also indicated a statistically lower number of login errors and login times for the passwords composed only of letters. However, the minimally restricted passwords were significantly weaker than the passwords that included a number and a special character.

Komanduri, Shay, Kelley, Mazurek, Bauer and Christin [12] compared passwords created by 5,000 participants, each assigned to one of five conditions across two sessions. The minimum length of these passwords ranged from at least 8 characters to 16 characters. Depending on the password condition, these passwords were either composed of only letters of varying lengths or included at least one character from each of the four character sets of a standard US keyboard. The results from the study indicated that participants took significantly more attempts to create restricted passwords of at least 8 characters than unrestricted passwords of varying character length. Compared to the shorter unrestricted passwords, the participants rated restricted passwords as significantly more difficult to recollect followed by the longer unrestricted passwords. Approximately 25 percent of the participants completely failed to create acceptable passwords in the restricted condition. However, the completion failure rates for participants in the other conditions were significantly lower: all under 19 percent.

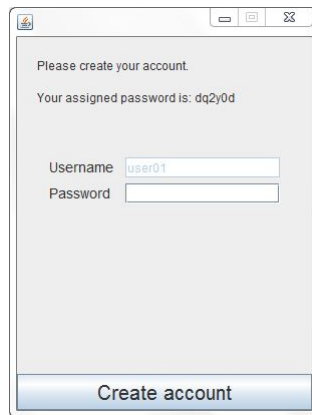
### 3 Method

This study compared the usability of three types of text-based passwords of approximately equal minimum security:

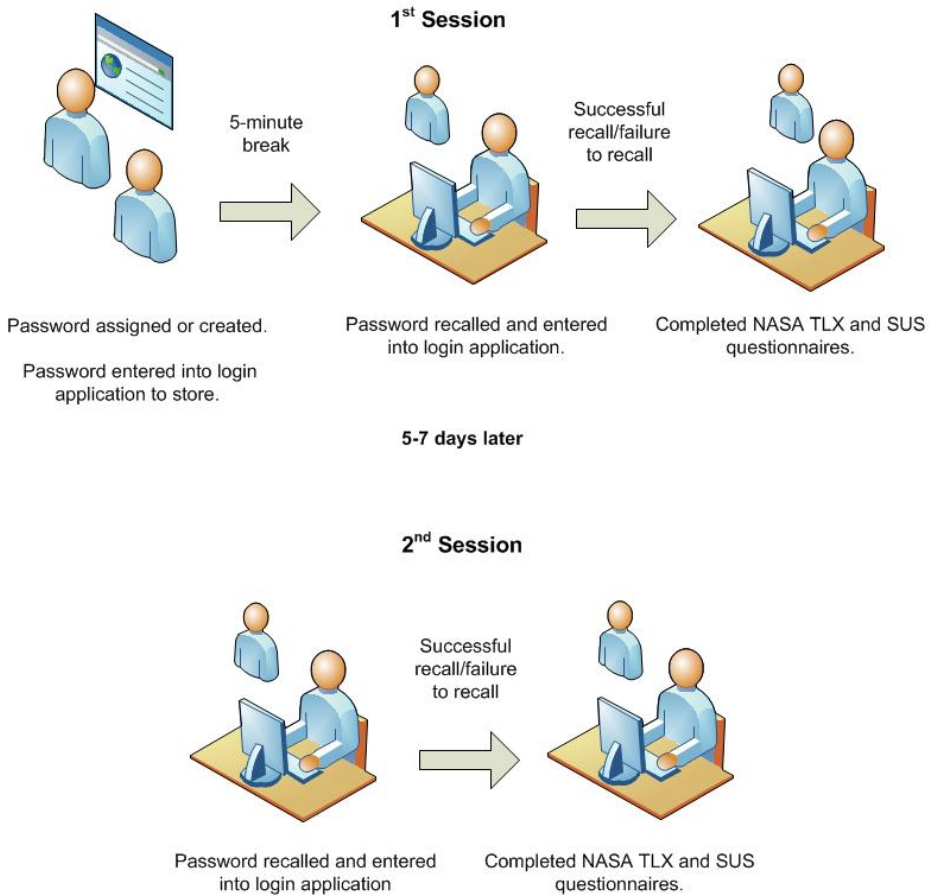
- 1: An assigned 6-character system-generated password selected randomly from any of the 36 alphanumeric characters available on a standard QWERTY keyboard.
- 2: A user-generated password of at least 8 characters, with at least one lower case letter, one upper case letter, one number and one special character. This password must also pass a dictionary check.
- 3: A user-generated password of at least 16 characters with no additional restrictions. This password must also pass a dictionary check.

#### 3.1 Study Participants

The study involved 54 participants, equally divided into three groups, with 18 in each password policy condition. The study took place over two sessions, with a period of 5-7 days in between them. In the first session, depending on the password policy condition, the participants were either assigned or they created a password and entered it into the password application (see Figure 1). If the password entered by a participant was entered incorrectly or failed to comply with the password composition policy, the application prompted the participant to re-enter a password. The participants were then asked to recall their passwords in the same session and after 5-7 days in the second session. The NASA task load indices [13] and the System Usability Scale (SUS) questionnaires [14] were administered at the end of each task: 1st session creation, 1st session recall and 2nd session recall (see Figure 2).

A screenshot of a web application window titled "Please create your account." The window has a light gray background and a blue border. At the top, it says "Please create your account." and "Your assigned password is: dq2y0d". Below this, there are two input fields: "Username" with the value "user01" and "Password" which is empty. At the bottom, there is a blue button labeled "Create account".

**Fig. 1.** 6-character alphanumeric password creation



**Fig. 2.** Procedural flow for first and second sessions

### 3.2 Experimental Design

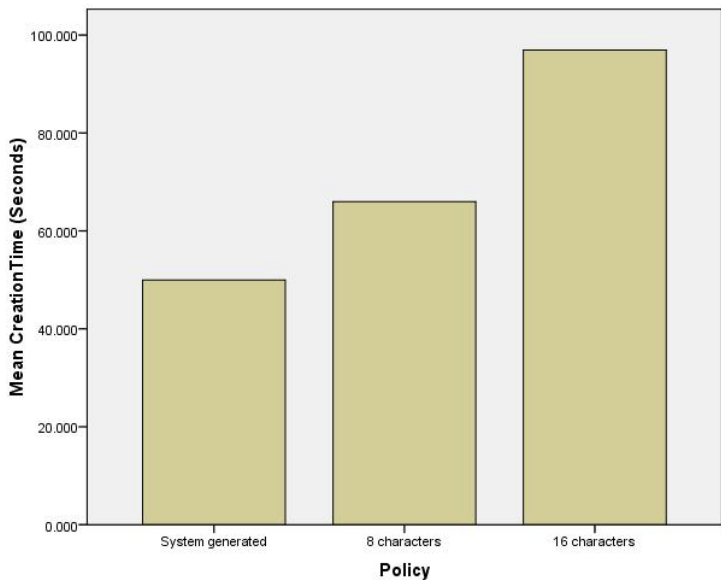
The three password policy conditions were compared with respect to the following dependent variables: the time taken to create the password account, the password creation error rate, the time taken to recall the password and recall error rate for both sessions, the number of unrecoverable passwords in the second session and the subjective ratings for the NASA task load indices and the SUS questionnaire for both sessions.

The experiment was considered to be a two-factor design. The first independent variable investigated the password composition scheme at three levels: the three types of password composition policies. The second independent variable of the study was task session. Although the main effect of task session was significant, this result was not a focus of the study.

## 4 Results

The results showed that it took less time to create an account with the system-generated password than with either of the two user-generated password conditions (see Figure 3). There were also significant differences between the password policy conditions for password creation error rate ( $p=0.002$ ) (see Figure 4) and the time taken to recall the passwords ( $p<0.001$ ) (see Figure 5). A post-hoc analysis revealed that the temporal demand index of the NASA-TLX questionnaire was higher for the 8-character user-generated passwords than for system-generated passwords ( $p=0.012$ ) (see Figure 6). There were no significant differences for recall error rate and unrecoverable passwords between the password policy conditions.

The results suggest that the overall performance of the 8-character user-generated password was weaker than that of the 16-character user-generated and 6-character system-generated passwords. A Pareto chart analysis of the comments made by participants (see Figure 7), as well as additional analysis of the user-generated passwords, suggest that participants were most familiar with the 8-character password policy condition. However, this familiarity did not translate into better memorability of 8-character passwords. The results suggest that the less familiar 6-character system-generated password and 16-character user-generated password composition schemes result in passwords that are at least as easy to recall, while imposing lower temporal demand.



**Fig. 3.** Mean password account creation time (seconds)

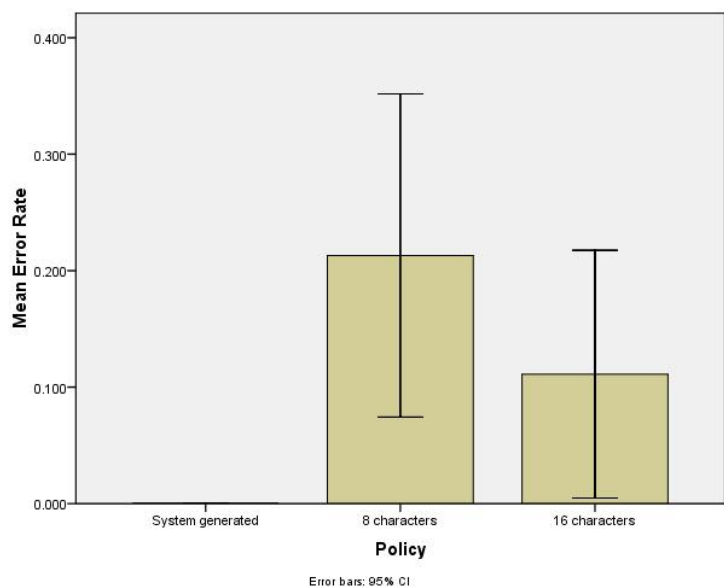


Fig. 4. Mean error rate during creation of password accounts

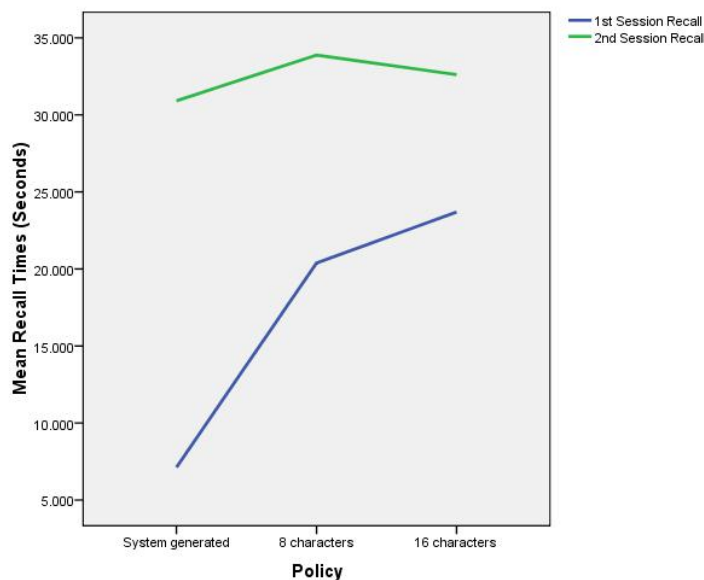


Fig. 5. Mean time taken to recall (seconds)

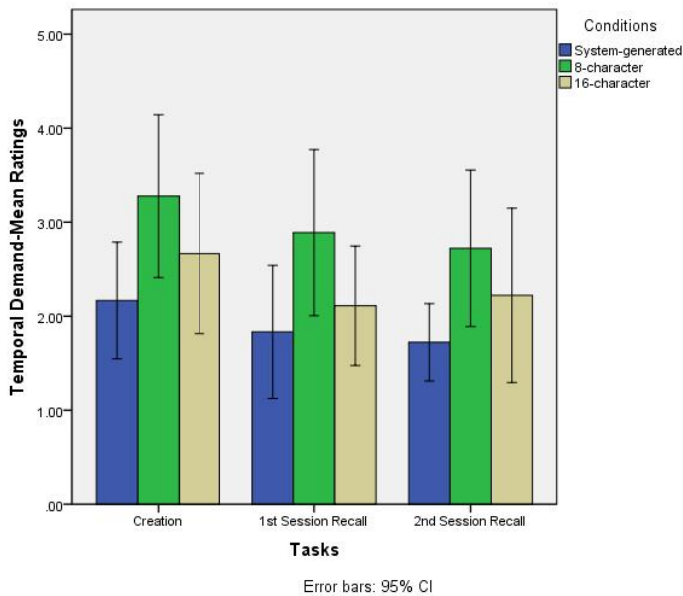


Fig. 6. Mean rating for temporal demand

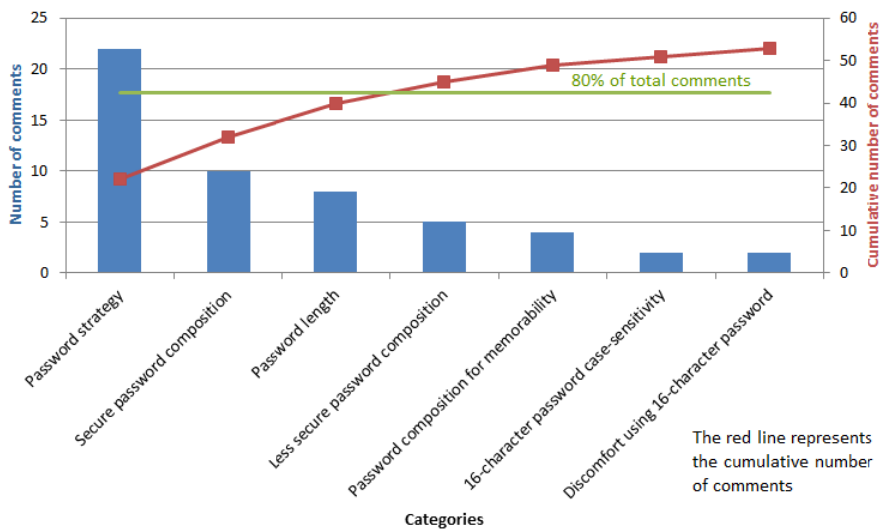


Fig. 7. Distribution of comment categories



## 5 Conclusions and Recommendations

This study compared the usability of three password conditions that assigned or helped users to generate passwords of approximately equal security, evaluating the trade-off between the length and the complexity of the passwords. The most important conclusion of this study is that the performance of the 8-character passwords was weaker than that of the system-generated passwords during the creation of password accounts and was weaker than the 16-character passwords in terms of long-term recall. Compliance with the restrictions associated with 8-character passwords strengthens security but creates passwords that are complex in composition. Thus, with the increase in applications requiring 8-character password accounts, a user may experience cognitive load when recalling a password from among competing passwords of similar composition. However, if 16-character passwords are created from a meaningful combination of preferably lower case letters, they may be more memorable than 8-character passwords subject to multiple restrictions.

Currently, the designers of password applications put most of the responsibility for creating a secure password on the users, forcing them to comply with a variety of restrictions. The complexity of such passwords may increase their security, but such security can also be achieved by increasing the minimum length of passwords and lowering the complexity of these passwords, thereby reducing the cognitive load on users. Thus, efforts should be taken to educate users on the trade-off between the length and complexity of user-generated passwords. A simpler and longer password can be as secure as a shorter but more complex one.

Designers should consider developing applications that aid users in creating longer but more meaningful passwords to reduce cognitive load. These applications could implement methods to produce 16-character passwords with meaningful combinations of letters, making these passwords more memorable to users. However, care should be taken by the designers to avoid explicitly restricting users to lower case letters only.

This study is a first step in exploring usable password conditions of approximately equal security. Below are suggestions for future research:

- Field studies involving participants belonging to a wider range of demographics.
- Studies involving the use of smartphones or tablets as password input devices.
- Studies on the effect of educating participants on the security of longer passwords composed of lower case letters.
- Studies involving a longer time period between creation and recall tasks to validate the results of the long-term recall of passwords across conditions.

## References

1. Adams, A., Sasse, M.A., Lunt, P.: Making passwords secure and usable. Paper Presented at the Proceedings of HCI on People and Computers XII, pp. 1–19 (1997), <http://portal.acm.org/citation.cfm?id=646684.702633> (retrieved from)
2. Adams, A., Sasse, M.A.: Users are not the enemy. *Commun. ACM* 42(12), 40–46 (1999), doi:<http://doi.acm.org/10.1145/322796.322806>

3. Zviran, M., Haga, W.J.: A comparison of password techniques for multilevel authentication mechanisms. *Computer Journal* 36(3), 227–237 (1993)
4. Conklin, A., Dietrich, G., Walz, D.: Password-based authentication: A system perspective. Paper Presented at the Proceedings of the 37th Annual Hawaii International Conference on System Sciences, pp. 10 (2004)
5. Proctor, R.W., Lien, M., Vu, K., Schultz, E.E., Salvendy, G.: Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers* 34(2), 163–169 (2002)
6. Forget, A., Chiasson, S., Biddle, R.: Helping users create better passwords: Is this the right approach?, Pittsburgh, Pennsylvania. Paper Presented at the Proceedings of the 3rd Symposium on Usable Privacy and Security, pp. 151–152 (2007),  
doi: <http://doi.acm.org/10.1145/1280680.1280703>
7. Allendoerfer, K., Pai, S.: Human factors considerations for passwords and other user identification techniques part 1: Field study, results and analysis (DOT/FAA/TC-05/20). Federal Aviation Administration William J. Hughes Technical Center, Atlantic City International Airport (2005)
8. Brostoff, S., Sasse, M.A.: Are passfaces more usable than passwords? A field trial investigation. Paper Presented at the Proceedings of HCI (2000),  
<http://hornbeam.cs.ucl.ac.uk/hcs/people/documents/Angela%20Publications/unsorted/hci2000.pdf>
9. Herley, C.: So long, and no thanks for the externalities: The rational rejection of security advice by users, September 8–11. Paper Presented at the New Security Paradigms Workshop 2009, NSPW 2009, pp. 133–144 (2009),  
<http://dx.doi.org/10.1145/1719030.1719050>
10. Burr, W.E.: National Institute of Standards & Technology. In: Burr, W.E., Dodson, D.F., Polk, W.T. (eds.) *Electronic authentication guideline* [electronic resource]: Recommendations of the National Institute of Standards and Technology/ (Version 1.0.2. ed.) U.S. Dept. of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, MD (2006)
11. Vu, K.L., Proctor, R.W., Bhargav-Spantzel, A., Tai, B., Cook, J., Schultz, E.E.: Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies* 65(8), 744–757 (2007),  
<http://dx.doi.org/10.1016/j.ijhcs.2007.03.007>
12. Komanduri, S., Shay, R., Kelley, P.G., Mazurek, M.L., Bauer, L., Christin, N., et al.: Of passwords and people: Measuring the effect of password-composition policies, Vancouver, BC, Canada. Paper Presented at the Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 2595–2604 (2011),  
doi: <http://doi.acm.org/10.1145/1978942.1979321>
13. Hart, S.: NASA-Task Load Index (NASA-TLX); 20 years later, October 16–20. Paper Presented at the 50th Annual Meeting of the Human Factors and Ergonomics Society, HFES 2006, pp. 904–908 (2006)
14. Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) *Usability Evaluation in Industry*. Taylor and Francis, London (1996)