Check for updates

HOW SOME ADVICE FAILS

William C. Hill

MCC Human Interface Laboratory
3500 West Balcones Center Drive
Austin, TX 78759

ABSTRACT

Video data for thirty-four cases of advice seeking, giving and following behavior at a graphical computer interface were analyzed in detail. The evidence indicated that clients followed prescriptive advice effectively and efficiently in slightly more than half the cases. For other cases, clients performed twice as many actions as needed in three times as much time and never reached prescribed states. A hypothesis that observed advice following difficulties were correlated with advice abstractness was not supported. Rather, it seems advice did not match well with clients' knowledge of the system in particular isolated details.

1. KEYWORDS

advising, collaboration, hidden operator.

2. INTRODUCTION

The presupposition that clients follow advice underpins belief in the utility of advice. The issue deserves investigation, because it might indeed be false. Mack, Lewis & Carroll (1983) [8] cited evidence that word processor users have difficulty following action prescriptions. Numerous studies have pointed out the difficulties of textual instruction following [3] [4] [5] [7]. At issue here is the question of advice effectiveness. Efficient following of good advice means effective advice; inefficient following of good advice renders advice ineffective.

During an earlier study [6] we observed that advice seekers did not always follow what otherwise appeared to be high-quality advice. We determined to look carefully at video records from this earlier study to estimate quantitatively how large a problem this was and to seek insight as to why it occurred.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1989 ACM 0-89791-301-9/89/0004-0085 1.50

3. METHOD

Existing data

Data available from the earlier study included video records of clients' work as they performed statistical tasks of their own creation in a role-play (as a *Consumer Reports* magazine data analyst). With at least a year of college-level statistics (so we might assume they had adequate statistics knowledge) and touch-typing skills, clients were individually tutored on the use of *Visual Statistics*, a direct manipulation interface and, in particular, how to seek and receive advice via its help key. *Visual Statistics* help-system worked as follows. After a press of the help key, clients could type English questions and send them to a hidden-operator advisory-system. After a few tens of seconds, English responses would appear on the clients' screens. Clients did not know advice was generated by a human expert behind the scene. Policies of limited initiative (i.e. only the client could initiate advisory exchanges) and licensed continuance (i.e. the advisor could continue exchanges with questions, requests and offers on a client-introduced topic) were enforced. These policies were deliberately chosen for the initial study to limit advisory phenomena to a simple case. Videotape records of the sessions were supplemented with mouse-move by mouse-move, mouse-click by mouse-click activity-script representations of client work and think-aloud protocols of advisor's advice justifications.

In order to maintain an ecologically plausible setting in the original study, we did not interrupt clients with think-aloud expectations or questions during the session. As a result, we lacked protocol evidence that would have permitted inferences with respect to the impact of individual pieces of *descriptive* advice (i.e., descriptions of the interface itself, rather than prescriptions of what to do at the interface) on clients' comprehension of the system. However, in the case of *prescriptive* advice, the video records and activity-scripts of clients' post-advice actions were available to characterize the relationships between action prescriptions offered by the advisor and what clients did with these prescriptions in terms of behavior at the interface. Requests for prescriptive

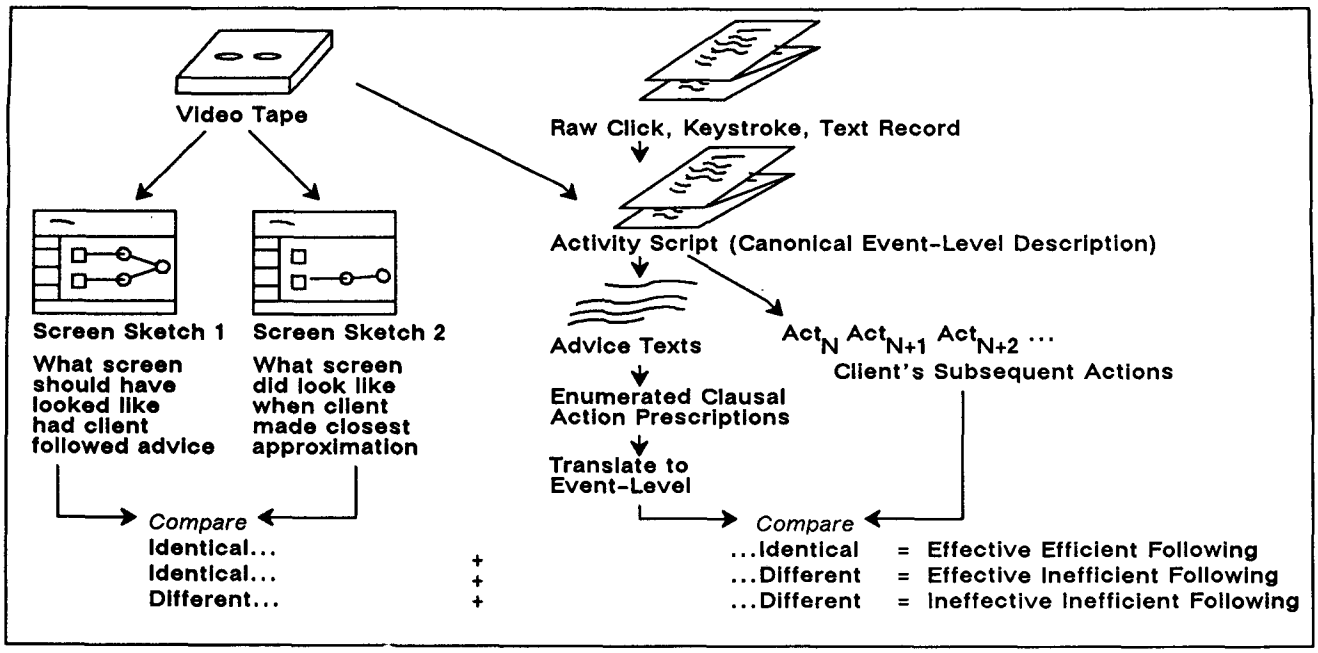

Figure 3-1: Diagram of Data Derivation

advice constituted 87% of observed requests, so we were able to measure advice effectiveness in a majority of advisory exchanges we observed.

New data from old

In order to clarify and quantify prescriptive advice-following behaviors, video and activity-scripts were used as follows to create new data. Consulting video tape as necessary, analysts performed four data-augmentation tasks for each prescriptive instance of advice. These were (1) sketch two interface screen-states, (2) deconstruct given advice into individual action prescriptions, (3) mark clients' subsequent actions sequentially, and (4) note times.

First, two sketches were produced from the video tape and activity scripts to capture screen-states in two crucial configurations: (1) the ideal screen -- what the screen would look like if advice were followed perfectly; and (2) what the screen looked like when the client's behavior brought the screen to approximate most closely the ideal screen prescribed in advice. Second, to deconstruct advice into individual action prescriptions, each grammatical clause in the advice that prescribed some action was underlined and isolated. For each such clausal action description, the sequence of *event-level actions*[1] necessary to perform the prescribed action was determined. This ideal sequence of event-level actions formed the basis for comparison with event-level actions actually performed by clients. Third, all the clients' actual event-level actions (those that occurred between the time the advice was received and the time that the client's screen most closely approximated the ideal screen-state implicitly prescribed in the advice) were collected for analysis. These collected client actions were compared with ideal actions as prescribed in advice and marked as either corresponding or not corresponding to prescribed actions. Fourth and last, times of closest approach to the ideal screen were noted. Figure 3-1 illustrates this derivation of new data from old.

These new data informed judgments of *presence* (i.e., were prescribed actions missing?), *sequence* (i.e., were prescribed actions completed in prescribed order?), *insertion of extraneous actions* (i.e., were unprescribed actions performed?), *faulty translations of action descriptions* (i.e., were prescribed actions performed incorrectly?), *timing* (i.e., how long did advice-following behaviors take?), *achievement* (i.e., what relation did clients' screen-states bear to ideal implicitly prescribed screen states?). Armed with these data concerning the relationship between prescribed actions and performed actions we could begin to characterize the advice-following difficulties we had previously observed in a less formal way.

Categorization of advice-following behavior sequences

Presence, sequence, extraneous action insertion, faulty action translation, timing and achievement judgments enabled categorization of advice-following behavior sequences as *effective efficient* following,

[1] A predetermined level for all interface actions existed to express possible action sequences at the interface.

Table 3-1: Summary of Advice Following

	Client Session		
	1	2	3
Total Advisory Episodes	13	20	6
Total Prescriptive Episodes	12	16	6
Exact Following	4	6	2
Exact except for Screen Rearrangement	3		
Exact except for Slip	1		
Effective, Efficient Following	8	6	2
Effective, Inefficient Following	2	1	
Inefficient, Ineffective Following	2	9	4

effective inefficient following, or *inefficient ineffective* following. How was this three-fold categorization operationalized? Two rules were used.

First, there was a rule to decide whether or not client actions were *effective* or *ineffective*. If an advice-following sequence achieved the prescribed state, it was categorized as effective advice-following; otherwise it was classified as ineffective following. Second, there was a rule to decide whether or not client actions were *efficient* or *inefficient*. If an advice-following sequence contained all prescribed actions in order without extraneous action insertion or faulty versions of actions, it was categorized as *efficient* following; otherwise it was classified as inefficient following. Exceptions to this rule were made for three instances of moving icons on the screen to form neater, more well-proportioned configurations and for one slip.[2] There were difficulties with the sequence of actions.

4. RESULTS

Table 3-1 summarizes the frequencies of advice-following categorizations. Selected clients participated in a total of 39 advisory episodes; 34 of these (87%) were prescriptive in nature. For effective following cases, the advice contained an average of 1.5 prescribed action description phrases that translated to an average of 5.9 event-level actions. Clients took an average of 85 seconds to follow advice in these cases. Extraneous actions introduced in effective, inefficient cases were inconsequential and took an average of five seconds each.[3]

For ineffective, inefficient following cases, the advice contained an average of 2.5 prescribed action descriptions, which translated to an average of 6.9 interface event-level actions. Using an average of 13.9 actions, clients took an average of 264 seconds to reach their closest approximation of ideal screen-state. Ideal screen states contained an average of 11.5 objects (icons and links). On the average, for ineffective, inefficient cases, clients' closest approximations lacked 1.3 necessary screen objects and contained 2.1 extraneous screen objects.

As expected, through practice and advice, client advice-following performance improved during the sessions, although this was not the central issue in this study. Figure 3-2 illustrates that production of extraneous actions was observed to decrease as a function of time (r=-.30). According to this observed correlation, at the session start, 55% of advice-following actions could be expected to be inefficient. An hour and a half later, at the end of the session, this average expected proportion fell to 18%.

The evidence indicates that clients followed prescriptive advice effectively and efficiently in slightly more than half the cases. For other cases, clients performed twice as many actions as needed in three times as much time without ever reaching prescribed states. What went wrong in these cases?

[2] It did not seem appropriate to classify tidiness actions as extraneous. The one slip occurred when a client opened a menu, quickly moved the mouse cursor to the bottom item on the menu and then slipped off the menu edge automatically closing the menu. The client immediately opened the menu again, moved the mouse cursor to the last item, slowly this time, and selected it.

[3] Once, a client opened a wrong menu and then closed it (this was not a slip). Once a client swept a half completed link around the screen. Once a client highlighted an icon marginally associated with the next prescribed step.

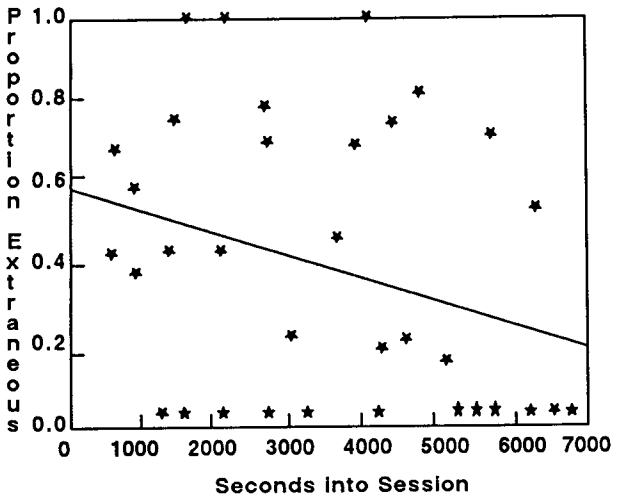

Figure 3-2: Scattergram of Proportion Extraneous by Seconds into Session All Advice-following Cases

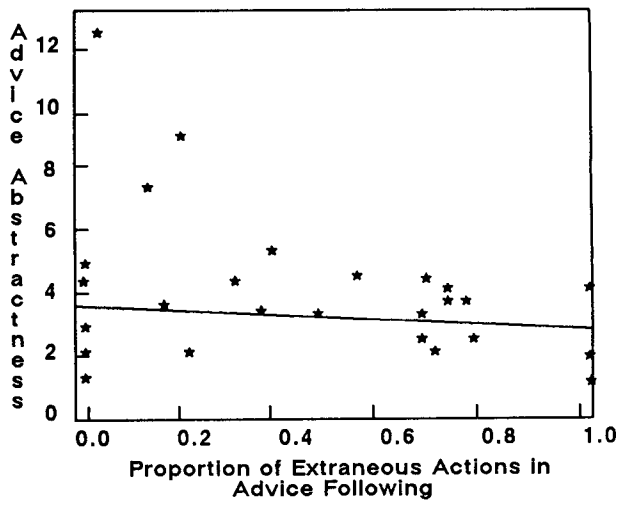

Figure 3-3: Scattergram of Proportion Extraneous by Advice Abstractness All Advice-following Cases

Perhaps the advisor was a poor advisor. The answer to this objection is "no". The advisor was an expert and master teacher. However, by no means does domain expertise and master teaching ability imply perfection, so perhaps the advisor advised poorly on some occasions. Perhaps the interaction possibilities, enabled by typed English exchanges with limited initiative and restricted continuances, were not rich enough to mediate assistance effectively. This seems probable since typed exchanges are less facile than spoken conversations.

One might hypothesize that difficulties were due to abstractness of advice, as measured by the ratio of the number of action prescriptions in the advice to the number of interface event-level actions into which they translated. The evidence does not support this view. Considering all cases of advice-following, there is a slight negative correlation (r=-0.16) between abstractness in advice and proportion of clients' actions that were extraneous with respect to the prescriptions (see Figure 3-3).

Failed prerequisite-presuppositions

Detailed analyses of advice-following, action by action, in relation to individual action prescriptions in the relevant advice revealed a pattern of *failed prerequisite-presuppositions* during advice generation. Records of the advisor's think-aloud advice justification stories confirmed these failures. Prescriptive advice failed when it violated Bloom's *mastery learning* thesis regarding the importance of prerequisites [1]. The advisor presumed the client knew something that the client did not actually know. In these failure cases, a client's lack of one or two small but critical details hindered his or her advice-following. Table 3-2 exhibits the failed prerequisite-presuppositions made by the advisor for all episodes that resulted in ineffective, inefficient following. Note that for Client 2 Episodes 3,4 the advisor continued to produce advice in accordance with the failed presupposition that the client knew that ports can have only one link. For Client 2 Episodes 5,6,7, the advisor continued in the failed presupposition that the client knew what was meant by "data column name". For Client 3 Episodes 2 and 4, the advisor continued in the failed presupposition that the client knew that to link to an icon, it must have been created previously. These observations suggest that an expert human advisor does not marshal all relevant available behavioral evidence to reformulate failed prerequisite-presuppositions while performing the time-critical task of advice production.

5. DISCUSSION: IMPLICATIONS FOR SYSTEM CONSTRUCTION

For advisory-system construction, three significant implications stem from poor advice-following behaviors and advice givers' failed prerequisite-presuppositions. First, if, as the results of this study suggest, roughly half of the requests for plan-type information end in deficient following behavior, the issue of licensed interaction style is a major one for system architecture. Second, mechanizations of advice generation suffer from greater but different limitations than human cognition. So, where a human advisor failed to take into account mounting client behavioral evidence, a machine programmed to do so would not fail (although its evidential reasoning might be very weak in other ways). This points out one small niche where mechanically produced advice might have a more consistently salient quality than advice produced within the cognitive limits of human expertise and memory load. Third, and most importantly, one can never be sure ahead of time that the presuppositions which ground advice are correct. They might always be wrong. Any advisory-system built with aspirations of actually helping more than half the time must be

Table 3-2: Advisor's Failed Prerequisite-Presuppositions Ineffective, Inefficient Advice-Following Episodes

- Client 1, Episode 1: The advisor presumed the client had the knowledge necessary to follow the scratchpad method of dataset copying.
- Client 1, Episode 11: The advisor presumed the client would perceive as critical a space character in an expression.
- Client 2, Episode 3: The advisor presumed that the client knew icon ports can have only one link attached.
- Client 2, Episode 4: Again the adviser presumed the client knew that an icon port can have only one link attached.
- Client 2, Episodes 5,6,7: The advisor presumed the client knew what a data column name was.
- Client 2, Episode 11: The advisor presumed the client knew what an "appropriate dataset" was.
- Client 2, Episode 13: The advisor presumed the client knew what ports on the select columns icons should be linked to the constant, "price" the "appropriate dataset".
- Client 2, Episode 14: The advisor continued to presume that the client knew that a port can have only one link (by this time, the advisor ha observed two instances indicating the client didn't know this).
- Client 2, Episode 19: The advisor presumed the client knew that one must first create a procedure before linking to it.
- Client 3, Episode 1: The advisor presumed the client knew what a data column name was.
- Client 3, Episode 2: The advisor presumed the client knew that one must first create a procedure before linking to it.
- Client 3, Episode 4: Again, the advisor presumed the client knew that one must first create a procedure before linking to it.
- Client 3, Episode 5: The advisor presumed the client properly perceived relevant aspects of a given constant expression.

constructed in a way to take this fact seriously.[4] Either the domain of advice must be engineered so that the system's rationale can be based on immutably correct presuppositions, or there should be some kind of *continuance* from the recognition of a failed prerequisite-presupposition.[5] The first option limits tractable domains to toy problems, so we are left with the second option.

It comes as no surprise to find that an advisor makes faulty presuppositions. How is it that face to face advisory exchanges deal with failed prerequisite-presuppositions in ways that our experiment forbade? Face to face advice overcomes such cognitive limitations through conversational *repair* [10], the iterative process of presupposition refinement. Realistic man/machine interactions for iterative presupposition refinement are at the frontier of human computer interaction research. A more feasible and immediate approach leaves the recognition of unsatisfactory advice (due to one or more failed prerequisite presuppositions) to clients by providing them with meaningful, easily accessible continuances from the current advice, such as elaborations of action descriptions, definitions of terms in advice, deictic references for terms, demonstrations, and examples. Such continuances would be first class pieces of advice themselves with their own continuances. The Intelligent User's Assistance Group at MCC's Human Interface Laboratory has constructed an advisory-system that incorporates both these approaches [9]. Empirical evaluations are underway.[6]

[4]This claim is consistent with Brown and Newman (1985) [2] advocacy for "design for the management of trouble".

[5]This recognition work might be performed either by the client or the machine. In the client case, interaction techniques must be made available for the client to express the recognition of a faulty presupposition. In the machine case, the logic that produces the communication also spawns "detective" processes to watch for the failure of presuppositions that ground the produced communication.

[6]Jean McKendree of MCC has begun an empirical study of the efficacy of two versions of this advisor system, one of which incorporated in part this continuance approach. Loren Terveen of the University of Texas and MCC is investigating feasible man-machine conversational patterns to cope with failed presuppositions in general.

6. ACKNOWLEDGMENTS

Thanks go to Jim Miller, Mike Masson, Tim McCandless, Loren Terveen, and Jean McKendree for numerous helpful discussions, Jim Hollan, Elaine Rich, and Peter Polson for reading earlier drafts of this paper and Susan Cook and Maria Nasr for assistance in analyzing data.

References:

[1] Bloom, B. *Learning for Mastery.* Durham, North Carolina: Regional Educational Laboratory for the Carolinas and Virginia, 1968.

[2] Brown, J. S., & Newman, S. Issues in cognitive and social ergonomics: from our house to Bauhaus. *Human-Computer Interaction* 1:359-391, 1985.

[3] Dixon, P. Plans and written directions for complex tasks. *Journal of Verbal Learning and Behavior* 21:70-84, 1982.

[4] Dixon, P. The structure of mental plans for following directions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13(1):18-26, 1987.

[5] Grosz, B. Focusing and descriptions in natural language dialogues. *Elements of Discourse Understanding.* Cambridge Unversity Press, 1981.

[6] Hill, W. C., & Miller, J. R. *Justified Advice: A Semi-Naturalistic Study of Advisory Strategies.* Technical Report, Austin, Texas: Microelectronics and Computer Technology Corp., 1988.

[7] Lynch M., Livingston, E., and Garfinkel, H. Temporal order in laboratory work. *Science Observed.* London: Sage, 1983.

[8] Mack, R. L., Lewis, C. H., & Carroll, J.M. Learning to use word processors: Problems and prospects. *ACM Transactions on Office Information Systems* 1:254-271, 1983.

[9] Miler, J. R., Hill, W. C., McCandless, T., McKendree, Jr., Terveen, L., & Zaback, J. *IDEA 1.0: Interactive Development Environment for Advice.* Technical Report ACA-HI-377-87-P, Microelectronics and Computer Technology Corp., Austin, Texas, 1987.

[10] Suchman, L. *Plans and situated actions: The problem of human-machine communication.* Technical Report ISL-6, Xerox PARC, Palo Alto, California, 1985.