

The Micro-structure of Use of Help

David G. Novick, Oscar D. Andrade, and Nathaniel Bean

Department of Computer Science

The University of Texas at El Paso

El Paso, TX 79968

+1 915-747-5480

novick@utep.edu, odandrade@miners.utep.edu, nhbean@gmail.com

ABSTRACT

To see whether, from the user's standpoint, muddling through with trial-and-error exploration of the interface actually works as well as consulting help, we analyzed video recordings of 14 subjects using Microsoft Publisher. We segmented their attempts at accomplishing four tasks into episodes, lasting from a few seconds to a few minutes, using four problem-solving approaches: help, recall, and trial-and-error. The subjects' overall rates of success with trial-and-error were higher than with help. We present trends of use of the approaches across the four tasks, review the apparent causes of success and failure when using the approaches, and develop an affordance-knowledge model that helps to explain people's preference for muddling through over using help. We conclude by discussing the model's implications for developers of systems and writers of help.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology, training, help, and documentation.*

General Terms

Documentation, Human Factors, Measurement

Keywords

Help systems, documentation, affordance

INTRODUCTION

Repeated studies of the use of computer help (e.g., [4],[7],[8],[9]) have shown that users of computer applications tend to solve problems through means other than help. Users typically “muddle through” by navigating the user interface with what often looks like a trial-and-error approach to finding solutions. It is possible to induce users to use help via a help-based tutorial, but the use of help appears not to provide users with much of an advantage, and the use of help tapers off consistently [10]. The reasons for these approaches to problem-solving, though, remain difficult to discern.

To make headway into understanding these patterns, we examined in detail people's actual use of help and other solutions in solving

problems. Looking at the entire context of use in which recourse to help may play a part, clues to people's successes and failures with an application should come from analysis of their actions as they attempt to complete tasks while guided by information from the interface and the help system. In previous work [10], we had been successful in inducing novice users of Microsoft Publisher to use help as they tried to accomplish tasks with the application, even if, as it turned out, those using help more did not apparently perform better on the tasks than those who used help less. The audiovisual recordings from that experiment provided a ready corpus for looking at what people did in their use of the application and its help system that led to success or failure. This analysis, in turn, can provide a start at guiding design of help systems (and interfaces) that more effectively aid people as they use an application.

The key questions we seek to answer in this study thus include:

- Why do users of computer applications decreasingly use help, even as they encounter problems not covered in their training?
- Do people using computer applications find that trial-and-error approaches take less time than first consulting help?
- What are the actual causes of failure to accomplish task steps, whether or not people use the help provided by a computer application?

In a companion paper [2], we look at the transitions among the ways that users try to solve problems; this could be termed a “macro” view of the relationship between help and other ways of solving problems. In this paper, we take a “micro” view, looking at what happens within individual episodes that use a single problem-solving approach. We seek to see whether, from the user's standpoint, muddling through with trial-and-error exploration of the user interface actually works as well as consulting and following help. We also begin to explicate what it means to “muddle through” by describing a range of problem-solving behaviors that run from seemingly random exploration of the interface to recalling exactly how to do something.

In this paper, then, we describe the study's methodology, provide examples of kinds of problem-solving by users, present the study's principal results, and discuss the implications of these results.

1. METHODOLOGY

This study starts with the audiovisual recordings collected in [10]. The subjects were 22 administrative assistants at a large university, of whom 21 were professional staff and one was a student employee. These subjects were recruited because they had little or no experience with Microsoft Publisher. Subjects were not compensated for their participation, but they gained some training

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGDOC '09, October 7-9, 2009, Bloomington, IN, USA.

Copyright 2009 ACM 978-1-60558-559-8/09/10...\$10.00.

and experience with the application, potentially useful in their jobs, as a result of their participation. The participants' mean self-assessment on general proficiency with computers was 3.18 on a scale from 1 to 5, indicating that they considered themselves to be of average proficiency. The participants' mean self-assessment on specific proficiency with Publisher was 1.64 on a scale from 1 to 5, indicating that they considered themselves to have little or no proficiency with the application. While the results in [10] were based on the data from only the 8 subjects who used help more than negligibly in their sessions, this study is based on the data from 14 subjects, so that we could have a more complete look at problem-solving behaviors.

1.1 Experimental Tasks

Each subject read a tutorial or received equivalent training through directed use of Publisher's help system; this introduction lasted 15 minutes. We gave the subjects 40 minutes to complete four tasks, where ten minutes were allocated for each task. Subjects were asked to replicate in Publisher every part of a one-page document (the "reference design"), which was handed to them for each task.

Of the four tasks, tasks 1 and 3 were designed so that the reference design could be replicated using features covered in the training; tasks 2 and 4 depended on features not covered in the training. Figure 1 shows the reference designs for two of the tasks. The tasks and the features used to create them were

1. "Arrow sign." Change the appearance of a text box, add a picture, rotate a picture, and add a background color or color frame to an image
2. "Business card." Locate the correct template, change the organization logo, and add a decorative frame to a text box
3. "Golden apple advertisement." Divide a textbox into columns, recolor a picture, and change how text wraps around a picture
4. "Raise petition." Change the color of a picture to grayscale, and add a shadow effect to text

The tasks were designed to be relatively easy, permitting measurement of the amount of time taken to complete each task. In fact, the tasks proved to be more difficult for the subjects than expected. The subjects often used the entire time available and sometimes did not complete the tasks. While this was problematic for assessment of effectiveness of the kinds of training the subjects had received, the tasks did lead to a great deal of problem-solving by the subjects, which meant that we had a correspondingly large amount of data to analyze in this study.

1.2 Data Analysis

To begin our analysis of the subjects' behaviors, we segmented the subjects' attempts at accomplishing the tasks into episodes, which lasted from a few seconds to a few minutes. Each episode represented the use of a single problem-solving approach, classified as *help*, *recall*, and *trial-and-error*. In all, the 14 subjects completed 189 usable episodes totaling about 252 minutes. The segmentation and analysis of the episodes was conducted through viewing of high-definition recordings of the participants' computer screens. Figure 2 is a screenshot of a participant working on task 1.

In our approach to categorizing the episodes, a *help* episode started with the consultation of help and included the subsequent attempt(s) to use that help; a new *help* episode also started when the user began a search for a new term. A *recall* episode started when a user appeared to know already how to do something

because he or she went directly to the right sequence of actions. A *trial-and-error* episode covered everything else and was meant to reflect the user exploring the interface, looking at, trying, and assessing the effect of various actions.

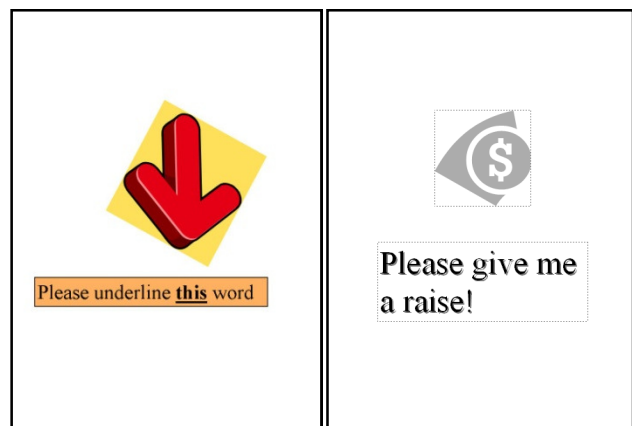


Figure 1. Reference designs for Reference design for task 1, "Arrow Sign," and task 4, "Raise Petition"

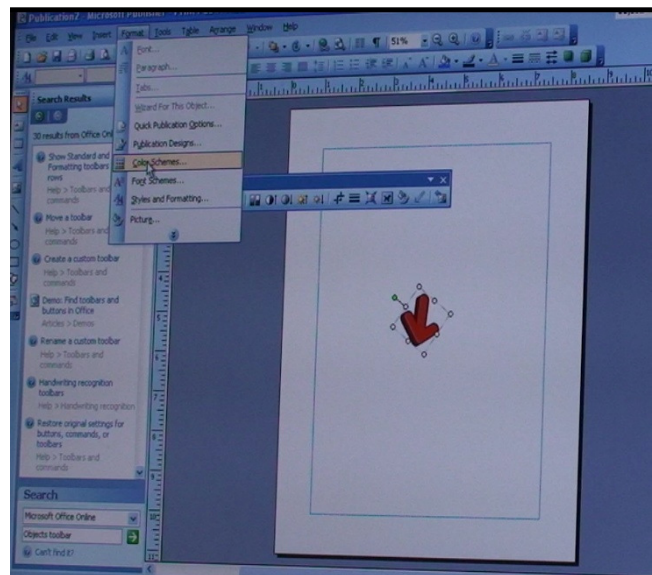


Figure 2. Screenshot during task 1.

For each episode, then, we classified the subject's approach (*help*, *recall*, or *trial and error*), noted whether the subject's actions in the episode were successful or not, described what they were supposed to accomplish, described what they actually did, noted the time spent in the episode, and commented on the apparent reasons for the episode's outcome. The initial codings were conducted by the second authors, checked by the first author, and recoded as required. All episodes in the *help* category were further reviewed to determine search terms used by the subject and the reasons why the search resulted (or did not result) in success in the task step. Likewise, all episodes in the *trial-and-error* category were further reviewed to determine the particular approach (explore interface, follow affordances) used by the subject and the reasons why the approach succeeded or not.

2. EXAMPLES

Before presenting the study's quantitative results, we first describe examples of successes and failures in subjects' use of the four approaches to problem-solving. These examples are intended to illustrate the level of detail of the analysis, to provide a fuller understanding of what the four categories represent, and to motivate the discussion of the results in Section 5.

2.1 Help Episodes

Our first two examples are cases where the subjects tried to use help to accomplish their tasks. In one case, the subject reached the part of a task where she intended to change the color of a clip-art image. She initially searched the Publisher help system for "change color," found and scanned an appropriate article, and then tried to navigate the interface by matching up key terms from the article with terms in the interface, instead of following the sequence of directions in the article. When this did not work, and after a trial-and-error sequence where she unsuccessfully tried various features, she returned to the help article that she had previously opened and scanned. This time she followed the instructions in the article, and quickly completed the task.

As was initially the case in our first example, the use of help did not always lead to success in the task. A different participant was also trying to recolor the clip-art image, tried to use help, and did not succeed. In this case, the subject navigated through the tutorial topics, found the "Recolor Picture" topic, and read the instructions. She then selected the image, hovered over the picture toolbar buttons to look for the *format picture* tool, selected "color," and closed the tool after seeing only the four initial options presented by the interface ("Automatic," "Grayscale," "Black and White," and "Washout"). At this point, the subject hovered over more options but did not go to the *format picture* button, and gave up. As at the start of the first case, the subject found the appropriate help but did not follow through by following the instructions. We believe that this behavior was caused by the subject's reliance on the interface's affordances [5], which may have had increased salience because they matched terms presented in help. We discuss this theory in detail in Sections 4 and 5.

2.2 Recall Episodes

In other episodes, the subject accomplished a task by recalling how to do a task step; the subject's recollection could come from prior experience with other Microsoft Office applications, the training in Publisher received in this study, experience accomplishing an earlier task in the study, or even seeing a function as an option while doing something else earlier. As an approach, recall was uniformly successful, by definition. For example, in one case a subject was supposed to insert and rotate a clip-art image of an arrow. These functions had been covered in the training. In the task, then, the subject selected the menu choices "Insert," "Picture," and "Clip Art," entered a query in the clip-art chooser for "arrow," navigated the list of results to select the red arrow and insert it and then rotated the image. In doing this, the subject followed this specific path of action and did not open other menus or try other functions.

2.3 Trial-and-Error Episodes

Subjects also tried a variety of other approaches that we grouped as *trial-and-error* ("T&E"), which included (1) apparently randomly exploring the interface's available functions (sometimes called exhaustive search) and (2) following the interface's

affordance, i.e., trying functions whose labels, either in text or in image, matched the subject's expectations of what was to be done. The T&E approaches were often successful. In one case, the subject was trying to change the color of a clip-art image to grayscale. The subject easily navigated to the *format picture* tool, and then spent time looking around in the window to identify the correct function. As one of the default options in the tool is "Grayscale," the subject was able to match up her understanding of what she was trying to accomplish with an appropriate term in the interface.

In other cases, however, trial and error was not a successful strategy. In one case, a subject was trying to insert an image of an arrow. She began by looking through interface for the function that would do this, but she was apparently unable to locate anything that she thought appropriate. However, the subject's exploration of the interface eventually led her to Publisher's "design gallery," which (a) appears in the "Insert" menu and (b) contains many images. The term "design gallery," in the context of the "Insert" menu, combined with an apparent trove of images, was easily sufficient affordance to draw the subject in that direction. Unfortunately, the correct way to do accomplish the task was to insert a clip-art image of an arrow, by selecting "Picture"/"Clip Art" from the "Insert" menu. The subject reached, in effect, a sort of local minimum from which she could not escape—the affordances holding the subject in the design gallery were stronger than the affordances that would have led the subject out of the design gallery toward the clip-art collection.

In some T&E episodes, the subjects arrived at shallow rather than deep solutions. The shallow cases were solutions where the steps were accomplished in a "surface" way that provided an approximation or the appearance of having done something rather than actually doing that thing such as, for example, arranging text in columns by inserting tabs rather than setting the number of columns. These "work-around" solutions typically conflicted with the application's model for handling text and images; the "deep" methods would have used the application's intended support for the function. That is, work-arounds were caused by the subject having an incomplete mental model of the task or the application. Nevertheless, the work-around approach almost always resulted in an apparently successful completion of that part of the task, even if the solution was not what an expert would have produced.

In one case, for example, the subject was trying to arrange the text in a text box so that she could fit a clip-art image between two sentences. Rather than format the image's text flow, the subject typed the first sentence, hit *enter* enough times to move the cursor all the way to the bottom of the text box, and then typed the second sentence. This made room for the image but was clearly not the best solution; it would have been better to format the image's flow, so that spacing between the image and the text would be precise and that the text spacing around the image would still be appropriate if, for example, the size of the image were changed.

3. RESULTS

Having identified, assessed and annotated the problem-solving episodes from the 14 subjects across the four tasks, we were able to obtain quantitative results for the relative effectiveness of the three approaches to problem-solving, for trends in the use of the approaches across the four tasks, and for identifying the key causes of success and failure in episodes where subjects used the four approaches. We found it helpful to distinguish two categories

of *help* episodes, referring to training and using search, and two categories of *T&E* episodes, exploring the interface and following affordances.

3.1 Relative Effectiveness of Approaches

The patterns of the subjects' use of help suggested that help was not proving to be significantly more useful than solving problems through trial and error. In fact, this turned out to be the case. Our analysis indicates that subjects' overall rates of success for *T&E* were higher than for *help*. Of the 124 episodes where subjects used *T&E*, 41.1% were successes; and of the 64 episodes where the subjects used *help*, 32.8% were successes (see Table 1). That this difference is not significant (chi-square, $p > 0.2$) likely reflects the limited number of data. Overall, though, the subjects appear to have experienced that the *T&E* approach was at least as likely as *help* to lead to success in accomplishing a task step.

Table 1. Relative success of trial-and-error and help

	Success	Failure	Success Rate
T&E	51	73	41.1%
Help	21	43	32.8%

The relative utility of *help* and *T&E* episodes could be affected by episode duration. If, for example, using the *T&E* approach took significantly longer than taking the *help* approach, then the apparent values of the approaches would differ. As shown in Table 2, time per success using *help* was less than time per success using *T&E*; the difference is statistically significant (chi-square, $p < 0.01$). Table 3 shows mean seconds per episode for successful and unsuccessful *help* and *T&E* episodes. In fact, the *T&E* episodes were shorter than the *help* episodes, and this difference is significant (chi-square, $p < 0.001$). The difference is largely due to the work-around subclass of *T&E* episodes, which tended to be briefer. The data in Table 3 for non-success also provide a view into the subjects' relative patience for the two approaches; subjects appeared to have been willing to spend more time on *T&E* than help before switching to another approach or giving up.

Table 2. Total time per success

	Total Time	Successes	Time/Success
T&E	9537 sec.	51	187 sec.
Help	4570 sec.	21	218 sec.

Table 3. Mean time per episode

	Success	Failure
T&E	79.5 sec.	80.3 sec.
Help	89.3 sec.	62.7 sec.

Another way of looking at the relative effectiveness of the two approaches is to compare time spent in successful and unsuccessful episodes. Of the 4,570 total seconds of *help* episodes, 41.1% of this time was spent in episodes ending in success; and of the 9,035 seconds of *T&E* episodes, 39.9% of this time was spent in episodes ending in success. These ratios are so close that it is unlikely that subjects could have perceived this difference.

In sum, the rates of success, total time per success, and mean time per successful episode all indicate that the *T&E* approach is more

effective than the *help* approach. The ratios of time in successful and unsuccessful episodes are nearly equal. But the mean success rates are different enough (32.8% and 41.1%) that that the subjects might have perceived a difference, even if strictly speaking that difference was not statistically significant. Indeed, the subjects used *T&E* in 106 episodes, compared to *help* in 64 episodes.

3.2 Trends across Tasks

Analysis of patterns of subjects' success and failures suggests that although users may initially view help as a more useful approach, they tend to shift to *T&E* experience. As shown in Figure 3, the subjects' use of help declined monotonically; this trend was surprising, because tasks 2 and 4 required skills not covered by the subjects' initial training.

Figure 3 also shows the subjects' use of the *T&E* and *recall* approaches across the four tasks. The spike in *T&E* episodes in task 3 is likely due to the characteristic of the task, which included dividing text into two columns and wrapping text around an image, both of which turned out to be relatively amenable to work-around techniques, such as creating columns by inserting tabs.

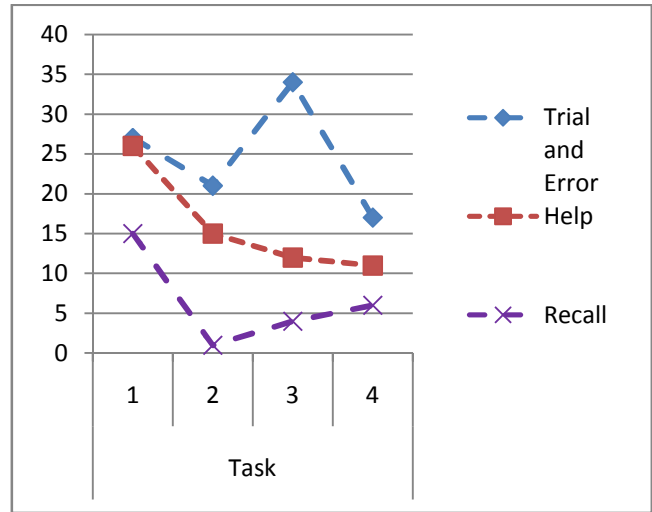


Figure 3. Number of episodes in which subjects used each problem-solving approach

Figure 3 shows the results by number of episodes. However, as shown in Table 4, the number of episodes varied across the tasks. It is possible that the number of episodes was lower in tasks 2 and 4 because subjects took much more time per episode in those tasks. Indeed, time per episode appears to account for about 95% of this variation.

Table 4. Number of episodes and time per episode, by task

	Task			
	1	2	3	4
Episodes	68	37	50	34
Time/Episode	67.1 sec.	91.8 sec.	90.0 sec.	98.1 sec.

Because the number of episodes varied across the tasks, the apparent trends in Figure 3 do not represent the full story. In Figure 4, then, we show the distribution of subjects' approaches in percentage terms. That is, Figure 4 presents the use of the three

approaches relative to the number of episodes in each of the four tasks. In this view, the *help* and *T&E* approaches have less of a tendency to decrease over time. The relative minima for *help* and *T&E* in task 3 are also less important than they appear, we believe, because of the idiosyncratic spike in work-arounds for that task. The variation in the use of recall may be due to two factors: (1) the immediacy of the training relevant to task 1, and (2) the subjects' growing experience with Publisher, as seen particularly in task 4.

Figure 5 shows the relative use of the problem-solving approaches in terms of time rather than number of episodes. In this view, the use of *help* remains relatively constant. The use of *recall* appears to fall off, but this trend is likely illusory because the *recall* episodes tend to be much shorter than the other kinds of episodes. The mean time of *recall* episodes was 29.8 seconds, compared with 76.9 seconds and 71.4 seconds for *T&E* and *help* episodes, respectively.

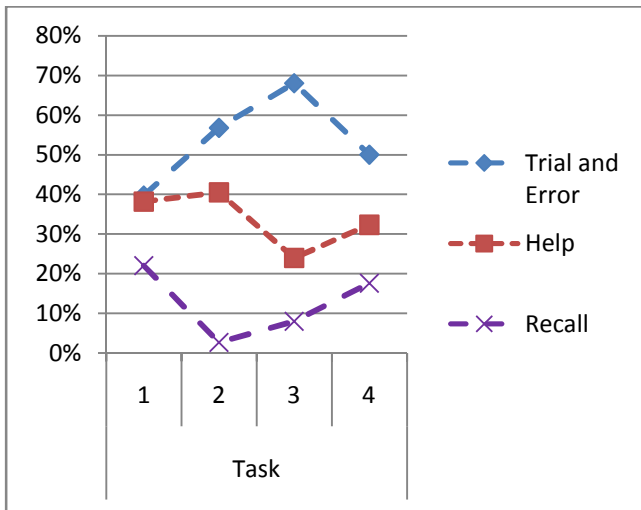


Figure 4. Relative use of problem-solving approaches, by number of episodes

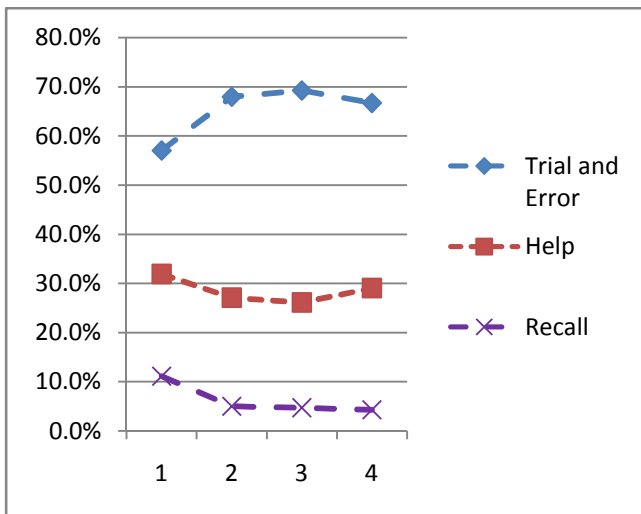


Figure 5. Relative use of problem-solving approaches, by total episode time

3.3 Causes of Success and Failure

So far, we have compared the relative effectiveness of the subjects' problem-solving strategies and looked at changes in their use of problem-solving approaches for possible insights into users' perception of the effectiveness of the approaches. We now turn from comparison of these approaches into analysis of the behaviors within the episodes, with a view to understanding why subjects' use of the approaches succeeded or failed. This analysis looked at every episode from all 14 subjects, with particular attention to the *help* and *T&E* episodes.

3.3.1 Help Episodes

In the recordings we annotated, the 14 subjects generated 64 episodes that we categorized as using the *help* approach to problem-solving. We were able to classify the episodes further as involving either (a) looking back at the training materials or (b) using search to find help.

The subjects had much greater success in reviewing the training materials than in searching for help; the subjects succeeded in accomplishing their task step in 53% of the training episodes and in 26% of the search episodes. The difference appears to be significant (chi-square, $p < 0.01$). Review of the episodes suggests that, significant or not, the results in the training episodes arose from these factors: (1) the training led to the subjects' having a more complete mental model of that feature of Publisher, so they were able to follow the instructions more easily; and (2) the training episodes necessarily involved tasks and application features with which the subjects had greater familiarity, and so conversely the search episodes involved tasks and application features that would pose greater problems for the subjects.

In the 43 search episodes that we were able to identify and classify, success and failure appeared to depend on the factors listed in Table 5. The factor most frequently affecting outcome of a search episode (14 episodes) is the match between the subject's expected term for a concept and the help system's use of a similar term. For example, the subject may want to "rotate" an image, and the help system uses the term "rotate" for that that concept; in this case, the subject easily finds the help they seek. But vocabulary mismatch likewise leads to episode failure. For example, the subject may want to change the "background" color of a clip-art image, but the system uses the term "fill" color for that concept; in this case, the vocabulary mismatch leads the subject toward help on changing the color behind an image instead of changing the color of the image itself.

Table 5. Factors affecting outcomes in search episodes

Number of Episodes	Factor
14	Vocabulary match or mismatch
11	Misunderstood model
7	Finds correct help, does not follow instructions
7	List of help topics did not afford right choice
4	Other

In seven additional episodes, the subject searched on plausible search terms, but the list of help topics displayed in response to the search did not apparently contain words that matched concepts expected by the subject; in these cases, the subject would either select a topic that was not actually relevant to her problem or

would give up on the search. For example, a subject, trying to add a border to a text box, searched the help system for “insert borders.” When presented with the list of help topics, she selected a topic about applying borders to tables, and so she was not able to apply the help. The help-topic episodes are similar to those where there was a vocabulary mismatch in the search terms, with the problem displaced from the search terms to the titles of the help topics.

Another factor that appears to affect outcomes in search episodes involves conceptual mismatches. In these eleven episodes, the subject did not understand the nature of elements of the application. For example, the subject may have thought that images are the same kind of object as text boxes, and thus tried to find and use inappropriate functions. This factor—misunderstood model—seemed to differ from the vocabulary and help-topic mismatch factors because it involves a misunderstanding of how the application operates; in the mismatch factors, the subject apparently knows how the application operates but does not know the correct names for some functions and objects.

Finally, in seven episodes the subject searched on appropriate search terms, found and opened an appropriate help topic, and then did not follow the instructions in the help topic. For example, a subject, intending to divide text into two columns, searched help for “column.” She selected and read an appropriate help topic “Create columns within a text box.” However, the subject did not follow the instructions at all (even though the instructions appeared clear to us). The subject then started unsuccessfully exploring the interface, looking at various menus, icons, and search results.

3.3.2 Trial-and-Error Episodes

In the recordings we annotated, the 14 subjects generated 124 episodes that we categorized as using the *trial-and-error* approach to problem-solving. We were able to classify the episodes further as involving either (a) exploring the interface or (b) following affordances. The following-affordance episodes included the work-around episodes.

By exploring the interface, we mean conducting what appears to be an unguided search for an appropriate function. This exploration could involve looking at menu items or icons seemingly at random, or hunting exhaustively through all menu items and icons. The principal characteristic of these episodes is that the subject appears to have little or no idea where to find a function appropriate to the task.

By following affordances, we mean looking at and trying functions whose terms or icons match, to some extent, the concept that the user seeks. As defined by Gaver [5], affordances are properties of the world that make possible some action to someone able to take that action; the affordances relevant here are perceptible affordances, in which there is perceptual information available for an existing affordance. In other words, the user of a computer interface sees that the interface enables the user to perform some intended function. In some cases, and our subjects encountered these, the perceptible affordance may be false: the interface appears to enable a function the user intends, but the actual function turns out to be something the user did not intend. Conversely, the interface may afford a function, but the affordance is not perceptible and is thus considered hidden. In other cases, the interface may afford a function that looks like it produces the right result but actually leads to a shallow result; these are the work-around episodes.

Not surprisingly, whether following deep or shallow methods, the subjects had much greater success in following affordances than in simply exploring the interface; the subjects succeeded in accomplishing their task step in 57% of the affordance episodes and in 17% of the exploration episodes. The difference appears to be significant (chi-square, $p < 0.01$).

In the 124 *T&E* episodes, looking at both the explore-the-interface episodes and the follow-affordance episodes, the subjects’ success and failure with task steps appeared to depend on the factor listed in Table 6.

Table 6. Factors affecting outcomes in trial-and-error episodes

Number of Episodes	Factor
55	Evident, hidden or false affordances
31	Vocabulary match or mismatch
38	Incomplete or mistaken mental model

In our analysis, the factor most frequently affecting the outcome in *T&E* episodes was the presence of evident, hidden or false affordances. In these 55 episodes, subjects succeeded when they were able to follow the cues or signposts in Publisher’s interface that signaled an appropriate function. For example, a subject trying to change an image to grayscale navigated to the “Format picture” tool, scanned the functions available in the tool (“Automatic,” “Grayscale,” “Black and White,” and “Washout”), and chose the “Grayscale” function. The subjects failed to complete task steps when this kind of affordance was missing or misleading. For example, some subjects, trying to insert a clip-art image, were led by the interface to the “Design Gallery,” which also had folders of images. Other subjects, also trying to insert an image, ended up in the “Graphics Manager.” Icons sometimes provided false affordances. For example, a subject trying to recolor a clip-art image selected the fill icon, apparently because it communicated the function of changing color; when use of this function changed the color of the background instead of the image, the subject undid the fill and began exploring fill effects.

In many of these cases, subjects could appropriately follow affordances, only to hit a dead end when the necessary cues ran out. That is, the subjects were succeeding in the task step by choosing the correct path in the interface, only to reach a point where the appropriate affordances were hidden. For example, subjects trying to recolor an image would correctly find and open the “Format picture” tool. But the tool’s default functions (“Automatic,” “Grayscale,” “Black and White,” and “Washout”) did not display the recolor function, which was in fact available in the tool. As a result, subjects would apparently think that they had navigated the interface incorrectly and abandon that approach to recoloring the image.

In another 31 *T&E* episodes, our analysis suggests that the factor affecting the episode’s outcome was match or mismatch of vocabulary. For example, a subject who was trying to add a border to a text box hovered over the “Text Box” item in the “Format” menu, but did not pursue this; apparently the subject did not connect her concept of border with the term “Object” in the menu that would have led her to functions for lines.

We attributed the outcome in an additional 38 *T&E* episodes to the subject’s mental model of the task or the application. Of these episodes, 22 were successes due to work-arounds. The other 16

episodes were failures, typically because the subject did not grasp how elements of Publisher worked or related to each other.

For example, a subject who was trying to change part of an image did not grasp that the parts of the image were grouped, so when the subject selected the part of the image that was her target, she also selected all of the other parts of the image, too. In another episode, the subject was trying to add a shadow to text. However, the subject ended up mired in the Word Art tool rather than finding the regular formatting tools for text. In this case, the evidence suggests that the subject’s model of how Publisher handled text attributes—or even that Publisher had text attributes—was incomplete or mistaken.

4. DISCUSSION

The quantitative data suggest that the *T&E* approach to problem-solving has greater utility takes less time than using *help*, although these differences are not all statistically significant. The differences in time and effectiveness do not seem big enough to explain the subjects’ nearly two-to-one (124 to 64) reliance on *T&E* over *help*. Why, then, did the subjects consistently rely on *T&E* more than they relied on *help*? In looking at answers to this question, we turn to our analysis of the qualitative data and to the theory of affordances.

4.1 Common Factors across Strategies

Despite their analytical usefulness, the distinctions between *recall*, *T&E*, and *help* episodes are fuzzy rather than crisp. First, as discussed in [2], subjects transitioned among approaches, often several times in attacking a single task step. Our review of the study’s recordings was replete with cases such as this one: The subject found the correct help topic, opened it, but did not follow the directions to complete the task. Instead the subject began exploring the interface for terms that might match terms in the help content. In a later episode, the subject returned to the same help topic and followed the topic’s instructions. Second, the different categories have key themes in common. Exploring the interface has much in common with following affordances—it’s just that the interface is not yet presenting cues that afford actions toward the subject’s goal. In work-arounds, the subjects are also following affordances, in that the subjects are using functions that are immediately afforded rather than using functions, perhaps with less obvious cues, that would accomplish the intended task consistent with the application’s model. Likewise, failures of the help system, such as displaying help topics that are opaque or misleading, can be seen as the effect of hidden or false affordances among the help topics. Even recall can be seen as a case of perfect affordance: the cues for navigating the interface and using functions are particularly clear because the subject already has knowledge of the cues’ actual meaning.

The connections among the problem-solving approaches can also be understood by considering the application’s communication. Users accomplishing tasks with the application are guided by a continuous range of information and cues, including communication from the application’s interface and communication from the help system. Most people seem to rely on the communication from the interface—the application’s affordances—and when faced with loss of affordance turn to strategies such as exploring the interface for cues or getting hints from the help system. Indeed, the application’s interface and help system both suffer from problems of hidden and false affordances: the application when, for example, users follow affordances to a dead end, and the help system when users select an inappropriate

help topic (e.g., about borders for tables when they need help with borders for text boxes).

4.2 Affordance-Knowledge Model

This analysis suggests a model of problem-strategies that has two dimensions, affordance and knowledge. The affordance dimension includes both the interface and the help system, and the knowledge dimension includes the user’s knowledge of both the application and the task. The model is shown in Figure 6.

Along the affordance dimension, the principal strategies identified in our study range from *recall*, based on perfect affordance born of experience, through *follow affordances* and its subclass *work around*, based on following cues to perform tasks through actual and simulated means, through *review training*, *explore* and *search for help*, which address lack of evident affordances.

Along the knowledge dimension, the strategies range from *recall*, which includes knowledge of both task and application, to *follow affordances*, which relies sufficient knowledge to recognize and interpret the interface’s cues, and *review training* which depends on recall of experience with the tutorial or with the interface, to *search for help*, which requires at least some minimal knowledge of what to search for, and *work around*, which uses a minimal amount of knowledge to achieve superficial results, through *explore*, which depends on little if any knowledge of either task or application.

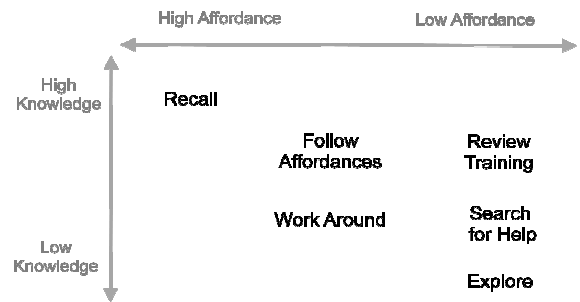


Figure 6. Affordance-knowledge model of problem-solving approaches

4.3 Preference for Trial and Error

The model’s affordance and knowledge dimensions provide some insight in addition to the suggestive evidence of the study’s data on relative effectiveness, into why people often prefer the *T&E* approach over *help*. This preference may be explained by four factors: the need for the user to have knowledge of the task or interface, the appearance of making progress toward completing a task, the conciseness of the interface relative to help, and the irrelevance of feature bloat.

Need for knowledge. Exploring the interface and searching help both result from loss of perceptible affordances, but searching requires at least some knowledge of the task or the application. First, the user must have some basis for forming a query. And second, the user must be able to interpret and choose among the help topics resulting from the search. In contrast, exploring does not require an *a priori* model of the task or the application; users may get the impression of learning as they go, even if their real knowledge of the interface or task does not actually improve.

Appearance of progress. Following affordances, which to an observer might seem like muddling through, may be appealing to users because they can perceive and follow cues or signposts with

apparent meaning. With the interface providing affordances, the intended result can always seem just around the corner, even if users follow the signposts into the trap of false affordance, and up to the point where they follow the signposts into the dead end of hidden affordance. From the users' standpoint, following affordances seems like making progress, while searching help is taking time in which they could have been making progress.

Conciseness. The interface's menus and dialog boxes usually present features and functions concisely, in a way that enables them to be scanned. This contrasts with help systems, which rely heavily on text, usually with linear flow, which tripped up some subjects in our study who appeared to scan the text for keywords to match up with cues in the interface rather than to read and follow the help as written.

Irrelevance of feature bloat. Users are likely to prefer exploring the interface over searching for help regardless of the number of features in an application. If the application's interface presents few features, then users can explore the interface in relatively short order. Yet if the interface presents many features, with many application having features amounting to bloat [6], then the interface presents all that many more possible cues and signposts. Many of these will necessarily be false affordances (e.g., Publisher's "Design Gallery" and "Graphics Manager"). But the users, who often cannot distinguish the false affordances from the appropriate ones, will have many more opportunities to make apparent progress.

4.4 Implications for Writers and Developers

If these four factors—need for knowledge, appearance of progress, conciseness, and irrelevance of feature bloat—are driving users' preferences for trial-and-error approaches over use of help, then developers and writers could likely provide better applications and help systems through these strategies:

- Work through the affordances for both application and its help system in tandem. This would involve studying the likely tasks of novice users, finding and fixing hidden and false affordances, and weeding and presenting help topics so that novice users with simple tasks do not end up in relatively advanced help topics (such as instructing how to find additional clip art on the Web when the user was trying to find the application's own clip-art resources).
- Ensure that the application's vocabulary matches users' expectations as much as possible. If subtle distinctions of vocabulary are unavoidable (such as, possibly, fill vs. background colors), use (a) roll-overs on the icons and menu choices to explain the functions concisely and (b) provide concept or terminology sections in the help system that enable users to distinguish terms that might confuse them.
- Use the help system to provide readers with accelerated understanding of the applications' model, not just key terms that will tempt readers to jump back into the interface and follow possibly false affordances.
- Develop interfaces [12] that are the minimal counterparts of the minimal manual [3]. If users are going to explore, regardless of the number of available features, minimal interfaces might help (a) limit false affordances and (b) reduce the time needed for exhaustive exploration.

- Consider ways of dynamically minimizing complexity in the help system [1]. This approach would enable users to tune the help system to their needs. The interface for this tuning would have to be extremely simple; otherwise the attempt to facilitate the users' problem-solving would require them to learn a new interface in addition to that for the application. Developing methods of automatically adapting levels of help to users' characteristics and needs would address this problem.

5. ACKNOWLEDGMENTS

This work was supported in part by an endowment from AT&T. Nigel Ward and the SIGDOC 2009 reviewers contributed many helpful comments that substantially improved this paper.

6. REFERENCES

- [1] Andrade, O., and Novick, D. (2008). Expressing help at appropriate levels, *Proceedings of SIGDOC 2008*, Lisbon, Portugal, September 22-25, 2008, 125-129.
- [2] Andrade, O., Bean, N., and Novick, D., (in press). The macro-structure of use of help, *Proceedings of SIGDOC 2009*, Bloomington, IN, October 5-7, 2009.
- [3] Carroll, J., Smith-Kerker, P., Ford, J., and Mazur-Rimet, S. (1987). The minimal manual. *Hum.-Comput. Interact.* 3(2), 123-153.
- [4] Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., and Shneiderman, B. (2004). Determining causes and severity of end-user frustration, *International Journal of Human-Computer Interaction*, 17(3), 333-356.
- [5] Gaver, W. (1991). Technology affordances, *Proceedings of CHI 91*, New Orleans, LA, April 27-May 2, 1991, 79-84.
- [6] Kaufman, L. and Weed, B. (1998). Too much of a good thing? Identifying and resolving bloat in the user interface. *SIGCHI Bull.* 30(4), 46-47.
- [7] Mendoza, V., and Novick, D. (2005). Usability over time, *Proceedings of SIGDOC 2005*, Coventry, UK, September 21-23, 2005, 151-158.
- [8] Novick, D., and Ward, K. (2006). Why don't people read the manual? *Proceedings of SIGDOC 2006*, Myrtle Beach, SC, October 18-20, 2006, 11-18.
- [9] Novick, D., Elizalde, E., and Bean, N. (2007). Toward a more accurate view of when and how people seek help with computer applications, *Proceedings of SIGDOC 2007*, El Paso, TX, October 22-24, 2007, 95-102.
- [10] Novick, D., Andrade, O., Bean, N., and Elizalde, E. (2008). Help-based tutorials, *Proceedings of SIGDOC 2008*, Lisbon, Portugal, September 22-25, 2008, 1-8.
- [11] Norman, D. Cognitive engineering (1986). In Norman, D., and Draper, S. (Eds.), *User-centered system design*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 31-61.
- [12] Vetere, F., and Howard, S. (2003). Layering a minimal interface, *Proceedings of IFIP INTERACT03*, Zurich, Switzerland, September 1-5, 2003, 415-422.