2014-05-01

Collocating Interface Objects: Zooming into Maps

May, J

10.1145/2556288.2557279
CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems
ACM

Collocating Interface Objects: Zooming into Maps

Jon May
School of Psychology
University of Plymouth
Plymouth, PL4 8AA, UK
jon.may@plymouth.ac.uk

Tim Gamble
Department of Psychology
University of Bath
Bath, BA2 7AY, UK
t.gamble@bath.ac.uk

ABSTRACT
May, Dean and Barnard [10] used a theoretically based model to argue that objects in a wide range of interfaces should be collocated following screen changes such as a zoom-in to detail. Many existing online maps do not follow this principle, but move a clicked point to the centre of the subsequent display, leaving the user looking at an unrelated location. This paper presents three experiments showing that collocating the point clicked on a map so that the detailed location appears in the place previously occupied by the overview location makes the map easier to use, reducing eye movements and interaction duration. We discuss the benefit of basing design principles on theoretical models so that they can be applied to novel situations, and so designers can infer when to use and not use them.

Author Keywords
Zooming; collocation; maps; cinematography; eye-tracking; cognitive models.

ACM Classification Keywords
H.5.2. User Interfaces: Theory and methods

INTRODUCTION
Interactive maps are everywhere nowadays, and map use has become routine [5]. Ramblers and drivers no longer have to battle with unwieldy sheets of paper, compromising breadth of representation against level of detail, with the crucial junction lost in the fold of the map, and the new road not even shown. Up to date cartographical information is now instantly accessible on computers, cars and handheld devices that are wirelessly connected to the internet, accurately positioned by GPS [9]. Relevant objects or businesses can be highlighted and irrelevant detail removed from the view, and the maps can be drawn at whatever level of detail the user requires for their current goal, zooming in from a low detail overview to successively more detailed local views of an area [3].

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

http://dx.doi.org/10.1145/2556288.2557279

Such flexibility comes at a cost, however, because the representation on the device cannot necessarily be the same as that on a paper map. A paper map is a designed representation of reality. It has been constructed by a skilled practitioner to take advantage of the map readers' attentional and perceptual processing, to allow them to selectively attend to certain classes of information in alternation [4]. Enormously successful, cartographic conventions have not developed overnight, and their evolution has relied upon developments in printing technology as well as the skill of the cartographer. In moving from a static, paper to dynamic, electronic representations, with far smaller screens and lower resolutions than are achievable in print, cartography faces a challenge. Allowing classes of information to be hidden or revealed adds interactional overheads to the design and use of a map that are absent when the map design takes advantage of selective attention. Changing between different scales of map as the user zooms in and out requires design decisions to be made about how successive views should be related. Designing an interactive map becomes a problem in human-computer interaction, prone to the same conflicting demands as other computer interface designs [5]. Technological change makes trade-offs between speed, capacity and mode of delivery difficult to anticipate. Fashions in implementation flourish and designers imitate more often than they innovate.

One thing that does not change, however, is the mental architecture of the map reader. It has previously been argued [2] that computer interface design should be a science as much as a craft, and should benefit by a principled understanding of the tasks users undertake, in order to capitalise upon the mental resources available for interaction. Early in the development of computer based maps, Moellering noted that 'when one begins to work with interactive display systems and particularly animations ... displays must be considered in a dynamic setting' [12]. Some researchers (for example, Young and Clanton [17]) have since advocated that interface designers learn from the dynamics of film editing, and should manage changes in their interfaces in the same way that film editors manipulated temporal and spatial jumps in films. May and Barnard [10] argued that rather than naively imitate cinematographic editing conventions, designers needed to understand why these conventions worked for films, and to create parallel changes in their interfaces based upon that understanding. Film editing had evolved as a craft skill, and its rules of thumb were situated within the domain of camera angles, lighting, and narrative, as sets of ‘dos and don’ts’, with no real account of why certain practices worked and were ‘filmic’, while others were ‘unfilmic’. While the work of researchers such as Kraft [6,7] had shown that conventions in film cutting did have a sound psychological basis, their domain specific nature made it hard for designers to know what it was they were to apply to their interfaces.

May and Barnard [10] presented a theoretical analysis of cinematography based upon Barnard’s Interacting Cognitive Subsystems (ICS) model of cognition [1], and made several extrapolations to interface design, including one scenario of a tourist information system combining large-scale and detailed views of a locale to help people find places of interest and plan routes. The problem posed was how to combine these two representations. In contrast to ingenious and novel combinations of fish-eye views, magnifying lenses and multiple simultaneous displays, it was argued that to be like a film the map interface should cut directly from an overview to a detailed view, without any swooping animation (as advocated in the context of data spaces by Mackinlay, Card and Robertson [8]).

The cognitive model of film watching that was proposed assumed that the main task of someone watching a film was to follow the narrative (in ICS terms, at the propositional and implicational levels of representation), and that any requirement imposed upon them to search the screen for relevant material (in ICS terms, at the object and propositional levels of representation) interfered with this task, detracting from their immersion in the story and making them aware of the surface detail at the expense of the story. Similarly, it was argued that in using a computer interface to carry out a task, any requirement for users to search the screen to relocate elements following a scene change would interfere with their main task, making the system less usable. Subsequent work on task-switching [13] makes a similar point, that interleaving two tasks adds cognitive overheads in terms of time and error compared to completing them in sequence.

In the context of map use, May & Barnard argued that the point that the user clicked on in the overview was where they were interested in for their task, and would be where they were still looking immediately after any screen change, and so the detailed view of the point clicked should be redisplayed in the same physical position on the screen: a principle they called collocation, as opposed to translation of the point of interest to a different spatial location. At that time, 1995, there were few such interactive maps readily available, but they noted the similarity between zooming-in with a map and with graphics packages. All but one of the graphics packages surveyed broke the collocation principle: they did a variety of things, often inconsistently, but most often they would translate the clicked element to the centre of the new view. The exception was Adobe Illustrator, the vector based partner of Adobe Photoshop (although in a subsequent release, it too joined the ‘move to centre’ camp).

One problem with the theoretically based principle of collocation was that there was no empirical evidence that filmmakers actually did make use of it. May, Dean and Barnard [11] gave a more detailed account of the cognitive model of film watching, and reported an eye-tracking study in which participants watched a commercially released film in its entirety. They measured the amount that observers moved their gaze location in the frames immediately after each of ten classes of cuts, and found that there was minimal eye movement in cuts where the cut zoomed-in to a detail, followed a moving actor or object, showed objects whose position was predictable, or showed the result or consequence of an action, or showed novel unexpected objects. In all of these circumstances, filmmakers had succeeded in collocating salient regions of the screen before and after cuts, so that the viewers did not have to move their eyes to locate something on the screen to comprehend the unfolding narrative. In five other classes of cut, large eye movements were found. Based on these findings, May *et al.* specifically argued that map interfaces should make use of collocation when zooming in and out to help map users rapidly locate the topic of their enquiry without having to search the screen for it. They also made recommendations about interactions where collocation would not be necessary, but where structural changes in the display might guide a user to the novel information, while leaving previous information visible to provide a context or history of the interaction. Such a situation, where translation would be better than collocation, might arise in the use of hierarchical databases, for example, where an object or name makes sense in the context of superordinate information and might be ambiguous or difficult to parse without this context remaining on the screen.

Despite this research, the translation convention was subsequently adopted for several computer-based maps, and in a recent evaluation of map interfaces, colocation was not even considered as an option [16]. There are clearly reasons why the translation convention was adopted, other than imitation or the re-use of open source code. Computationally, all that needs to be returned to the map server are the absolute co-ordinates of the clicked location and a scaling factor (either of the current or of the new view), as can be seen by inspecting the URLs, which generally include a latitude, longitude and level parameter. Without knowing where within the previous image the clicked location was, the obvious place to put it in the new view is in the centre. If the user is going to zoom in further, then it will remain at the centre. It could even be argued that such interfaces are consistent, which has become a watchword of design guidelines, and hence easy to learn and use.

However, translation also has disadvantages, from the point of view of the user. If they have been looking in one point on the screen, the place they are interested in is now not there, and they have to search for it. If the new representation is stylistically different to the original, as is often the case with maps of different scale, then it is not an easy task to re-orient oneself to the new view and to relocate the place name of interest. If one has not clicked on the actual place, but on or near its name, which is usually to one side of the place, then the actual location desired will not be in the centre of the screen, and if its name has been repositioned on the new view it may be anywhere within the screen, depending upon the scaling factors. If one is unfamiliar with a geographical region, as is often the case when using a map, each wrong place name one reads gives no help; serial search is all that is possible. On the other hand, translation has some advantages for the programmer, because to collocate the result with the original map, two additional parameters need to be returned to the server to indicate the relative location of the point of interest within the first map, so that it can re-occupy that position in the new map. In essence, the decision between the two modes of presentation appears to be a contest between ease of use and ease of programming.

From a psychological point of view, the only way that a move-to-centre translation algorithm could be of benefit to map readers rather than to map programmers, is if the widespread adoption of the convention has been learnt by the general public and so they expect to look in the centre of maps as a result of an interaction with a place on the periphery. This is the question that is set out to be answered with the three experiments reported in this paper. By contrasting an interface using collocation with the same design but using translation, it should be possible to see if the theoretically principled recommendations provided any advantage for users, or whether they had grown used to the convention and by knowing where to look would be able to interact as easily with the translated interfaces.

EXPERIMENT 1

Method

There were 26 participants in the study, all of whom were paid £5 for their participation. They were told that the study was investigating eye movements people made while zooming into maps, but not that either method of zooming was of particular interest.

The two interface techniques of collocation and translation were presented to the participants as different computer systems, both using maps. One was a traffic control system, in which their task was to locate traffic monitors; the other was a gas supply system and their task was to locate gas meters. The interface was written in RealBasic and run on a 2.7GHz G5 Mac using a 20” NEC Multisync 2080UX+ LCD monitor at 1024 x 768 pixel resolution, viewed by participants from a static operator’s chair with headrest (to minimise head movement) at a distance of approximately 75 cm. Each map was 297mm square, subtending approximately 22 degrees, with the centre of the screen approximately level with the participants’ eyes.

Thirteen scenarios were created, using Ordnance Survey maps of different cities in Britain (the maps were obtained from the Edina Digimaps service under an educational license). Each scenario began with a text message asking participants to locate a target in a specific named location within a city, e.g. ‘There is a traffic monitor located near Walkley in Sheffield. See if you can find it’. On clicking OK, the screen cleared and a start button was displayed centrally. When clicked, an overview map of 1:50,000 scale was displayed and the participant had to find and click on the specific local placename (e.g., Walkley). Following the click, a 1:25,000 scale map of the area clicked on was displayed, and the participants had to relocate and click on the placename, to be shown the final 1:12,500 map (produced by rescaling the OS 1:10,000 map). This final map had a target symbol (a black diamond 4mm high and wide) embedded within it, upon which the participant had to click to complete a trial. The target was placed in the area that the participant had been asked to search, but was not systematically located in relation to the place name on that view. If at any point the participant clicked on a region not containing the target area, they could ‘zoom out’ to the previous map using option-click.

Twelve of the scenarios were divided between two experimental blocks. In one block the maps were collocated such that the geographical location under the mouse cursor was identical before and after the click, notwithstanding the change in scale. In the other block the conventional translation approach was used, such that the geographical location clicked on the first map was placed in the centre of the second map (Figure 1). The remaining scenario was used as a practice trial before each block.

The traffic monitor and gas meter cover stories were balanced over experimental conditions, and the order of the conditions was also balanced over participants. The twelve scenarios were presented in a Latin square design over participants, such that after twelve participants each scenario had been used in each serial position within each condition.

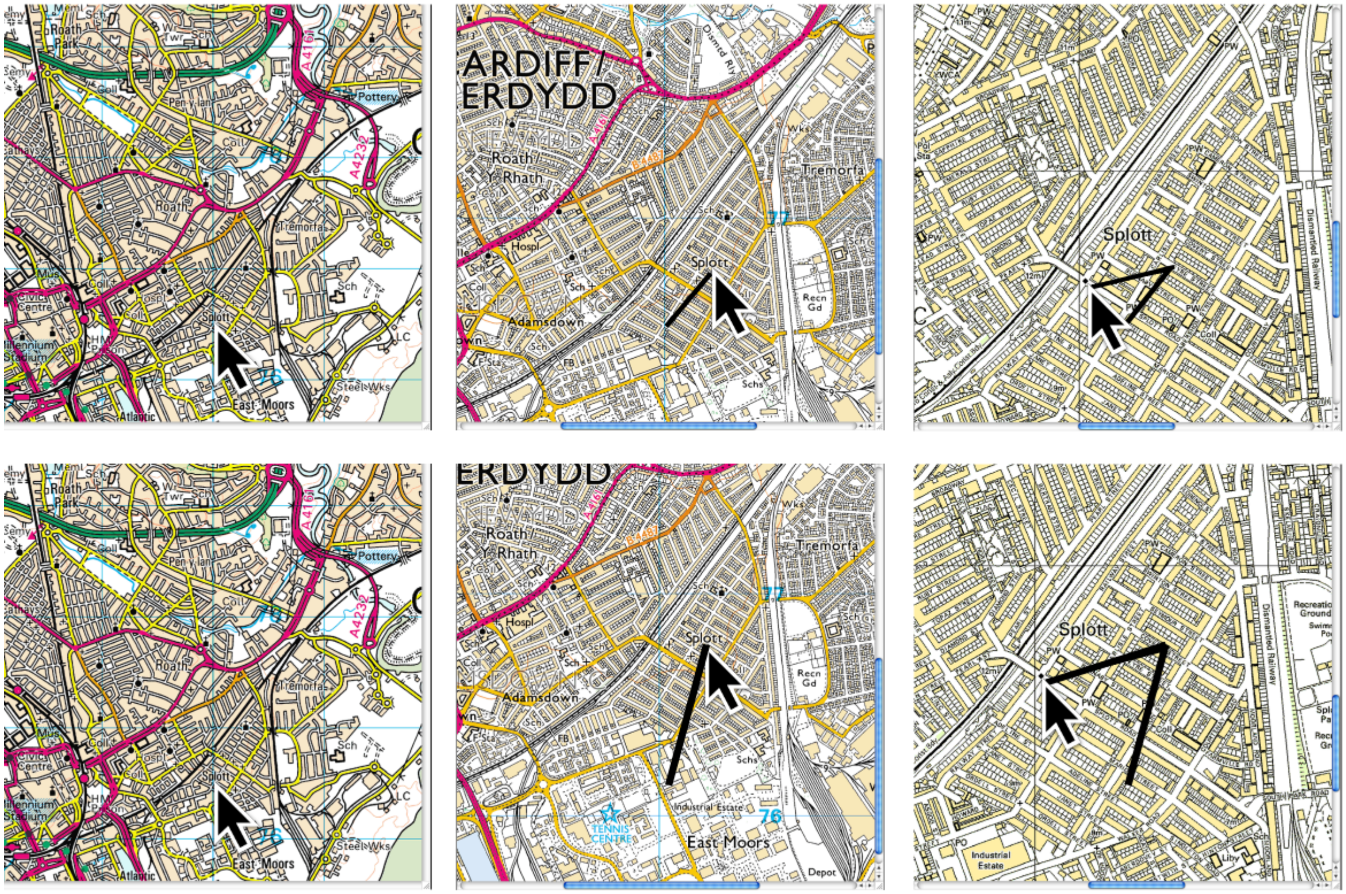

Figure 1: In the collocate condition (upper row) the geographical detail clicked on is positioned under the cursor on the new, higher scale map. In the translate condition (bottom row), the clicked detail is moved to the centre of the new map. In both conditions, the task is to *find* a place name in the first map (the Splott area of Cardiff), *relocate* it in the second map, and *select* a diamond shaped target symbol in the third map. The arrow cursors have been exaggerated for clarity, and indicate the point that is about to be clicked, and the lines show the path traversed. Note that the location of the place name relative to the street layout varies from map to map. Maps (c) Crown Copyright/database right 2008. An Ordnance Survey/EDINA supplied service.

Each participant was tested individually. Participants' eye movements were recorded using a SensoMotoric Instruments iView X eye tracker, which requires no head restraint or worn head position tracking device, thus making the situation as naturalistic as possible. The screen coordinates of each participants' gaze location were logged every 20 msec for the duration of the experiment. The presentation computer recorded reaction times for each experimental event.

Results

Reliable eye tracking data could not be obtained for four participants. Of the 22 remaining participants, ages ranged from 18y 9m to 27y 2m, and three were male. Eleven completed the study in each order (i.e., Collocate-Translate or Translate-Collocate). Raw eye tracking data was analysed using the BeGaze software, with parameters defined to detect fixations with a minimum duration of 80 msec and a maximum dispersion of one- sixth the screen height. This reduces the raw data to a series of fixations (where the pupil is relatively stationary), saccades (where the pupil is moving) and blinks (where the pupil is not detectable for several frames). If the software reported more than three blinks in a step, the data was discarded as unreliable, because the software does not distinguish between true blinks and episodes where the pupil cannot be detected due to tracking problems such as head movement. Trials in which the participant used the zoom-out function to step back were also discarded. Over the 264 trials from the 22 participants, data was rejected from 39 of the Find steps, 14 of the Relocate steps, and 18 of the Target steps (9% overall).

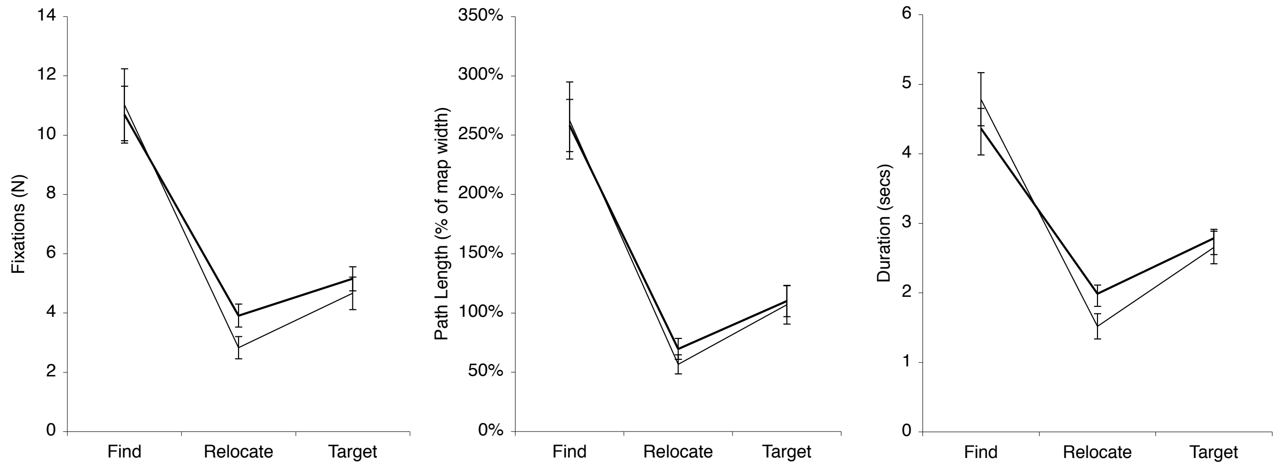

Figure 2: The two interfaces performed identically on the *find* and *target* steps, but the translate interface (bold line) required more gaze fixations, a longer search path, and took longer to complete on the critical *relocate* step, than did the collocated interface (narrow line). Bars indicate one standard error.

From the resulting analyses, the number of fixations, their total path length (expressed as a percentage of the width of the map), and the overall duration of each step in each trial was obtained (Figure 2). These three dependent variables were entered into a MANOVA with the factors of condition and task step. There was no overall effect of condition (Pillai's trace F(3,19)=0.69), but there was an effect of task step (Pillai's trace F(6,82)=13.5, p<.001, η_p^2=.497) and an interaction (Pillai's trace F(6,82)=2.60, p=.023, η_p^2=.160).

Univariate tests showed no overall effects of condition for any of the three measures (all Fs<1), but significant effects of task step for all three (Fixations F(2,42)=52.6, MSE=12.3, p<.001, η_p^2=.715; Path lengths F(2,42)=58.5, MSE=1.06, p<.001, η_p^2=.736; Durations F(2,42)=70.9, MSE=1.27, p<.001, η_p^2=.771), and interactions of condition with task step for Fixations (F(2,42)=3.60, MSE=5.64, p=.036, η_p^2=.146) and Durations (F(2,42)=5.04, MSE=0.85, p=.011, η_p^2=.194), with a non-significant interaction for Path lengths (F(2,42)=2.80, MSE=0.67, p=.07, η_p^2=.118).

To interpret the interactions, one-tailed paired t tests were then conducted between the two conditions in each task step. There were no differences between the two conditions in the Find step (Fixations t(21)=1.22; Path-length t(21)=1.05; Duration t(21)=1.72; all p>.10) or for the Target step (Fixations t(21)=1.17; Path-length t(21)=0.15; Duration t(21)=0.34; all p≥.10), but on the Relocate step the translate interface required more fixations (Col=2.76, Trans=3.89, t(21)=2.36, p=.014), longer path lengths (Col=65%, Trans=105%, t(21)=2.59, p=.009), and took longer to complete (Col=1.54s, Trans=2.00s, t(21)=2.22, p=.019).

Discussion

The two interfaces did not differ on the initial Find step, where participants were searching for the name for the first time, nor on the Target step, where they were searching for the diamond symbol, but they did differ on the crucial Relocate step. At this point in the task, participants had already found the name and clicked on it on the first map, and had to find and click it again. The conventional translation approach of presenting the clicked point in the centre of the new view required more and longer eye-movements and took participants longer to execute than the advocated principled approach of collocating the position on successive maps. This difference exists even though the translation approach is consistent, because the clicked point is always in the same physical screen position in the new view and so participants could simply make a single change in gaze direction. Instead, they are not able to make use of this consistency, having to search for the place name again.

The lack of benefit of collocation on the target step supports the proposed argument because the diamond was fixed on the map, and not collocated with the place name or the location clicked by the user. In practice, the locate step required visual search in both interfaces.

It could be argued that this version of a map-searching task does not give much benefit to the consistency of the translation approach because with only three levels of map to search through, there is only one critical trial upon which this consistency can be used: the transition from the second map (in which the clicked point has been translated to the centre) to the final map. The next experiment extends the task to a five level task, so that having selected a point on the first map, and having had it translated to the centre, subsequent clicks on maps two, three and four should also be close to the centre of the map, potentially making a translation approach more useful, and in practice, more like a collocation approach.

EXPERIMENT 2

Method

There were 25 participants in this experiment, and each was paid £10 for their participation. They were told that the study was investigating eye movements people made while navigating computer-based maps. The experiment was conducted using the same equipment as Experiment 1.

Twenty-four map-reading scenarios were created using Ordnance Survey maps of the Plymouth area, with which none of the participants were familiar. Each scenario began with a text message instructing participants to locate a local place name in a numbered region of Plymouth. Participants read this instruction aloud. On clicking OK, an 250mm square overview was displayed, based on a 1:50,000 scale map, upon which the numbers 1 to 12 had been superimposed in a clock face arrangement (Figure 3). After clicking on or near the relevant number, a twice as detailed view of a quarter of the original map was displayed. Clicking on this displayed a third level view, again twice as detailed, and so this time of 1/16th of the original map. Clicking on this presented a fourth (1/64th) and finally a fifth level (1/256th) view. The first three levels were all based on the same 1:50,000 scale map, level four on a 1:25,000 scale map, and level five on a 1:12,500 scale map (rescaled from the OS 1:10,000 map). When the target place name had been clicked in the fifth level map, a congratulation message was displayed. If at any time a click would have resulted in the place name being off-screen, an error message was displayed and the trial was restarted from the instruction message (this meant that the zoom-out function did not have to be made available).

The place names used as targets were spurious, being taken from a different region (Northumberland) and added to the maps so that while they were in roughly the same geographical location on each map, their actual position and typeface varied in the same way that real place names varied in typeface and position on the three original maps used. Participants undertook two sessions (approximately three months apart), each of 24 trials divided into two blocks. One block in each session used the Translate method of zooming in, so that the new view was centred upon the location represented by the point clicked in the previous level; and the other block used the Collocate method, so that location represented by the point clicked in the previous level was presented in the same relative location within the new map window.

Block order was balanced over sessions and participants, so that half experienced Translate first followed by Collocate in the first session, and then Collocate followed by Translate in the second session. The twenty four scenarios were balanced so that half were used in the translate condition in session one and the collocate condition in session two, while the others were used in the collocate condition in session one and the translate block in session two. Within each block, the order of trials was rotated over participants in a Latin square, so that each target appeared at least once at each position in the block.

Results

Three participants were unable to return for the second session, and eye-tracking data from one participant was too poor for analysis. This left 21 participants' data for analysis (6 males; age range 20-36 years, median 24 years). Of these 1008 trials, 48 (5.2%) were repeated due to errors, with one participant making 8 errors (17%); the next worst made 4 errors (8%). Errors were more frequent in the second session (n=29) than the first (n=19), the reverse of a practice effect, but were more likely near the start of a block (number of errors per four trials: 17; 10; 5; 7; 5; 4). They were also more frequent in the Translate condition (n= 30) than in the Collocate condition (n=18). A binomial test gives a probability of .056 of observing 18 outcomes or fewer out of 48.

For each of the five steps in each trial, three measures were computed as in the previous experiment: the number of fixations, the total path-length, and the duration of each step. For each participant, a mean for these measures was computed over the twelve trials in each block, resulting in a 2x2x5 repeated measures design with the factors of Session, Method, and Step.

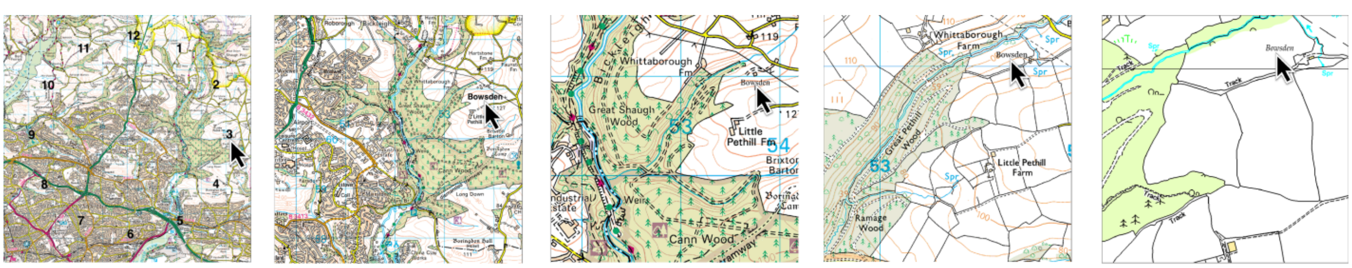

Figure 3: In the second experiment, participants had to locate a place name that had been added to five increasingly detailed maps such that its appearance and geographical position varied from map to map, as real place names do on different scale maps of the same region. In this example, the task is to locate Bowsden using the collocated interface. The cursor has been exaggerated for clarity. Maps (c) Crown Copyright/database right 2008. An Ordnance Survey/EDINA supplied service.

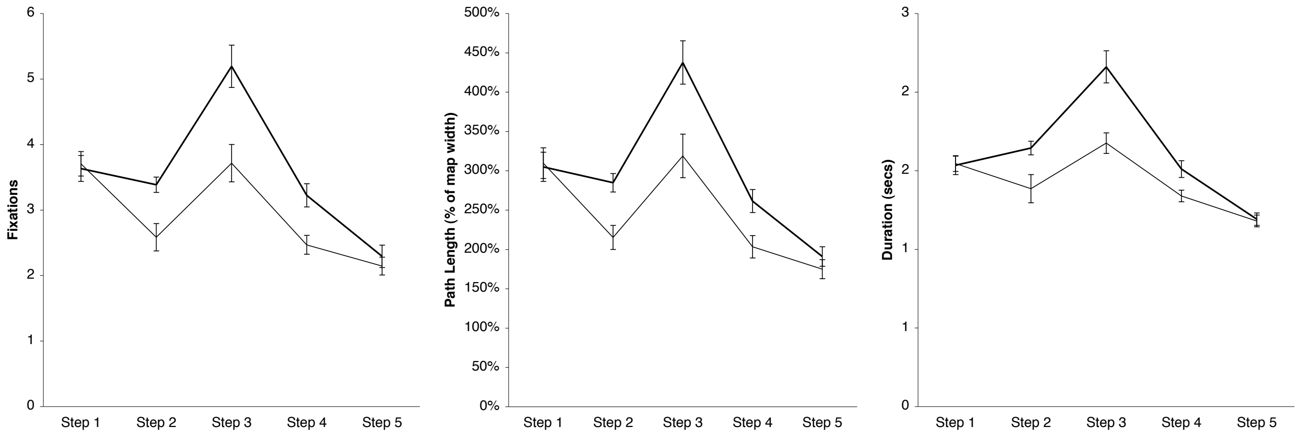

Figure 4: The Translate method (bold line) required a longer path length, more fixations, and took more time than the Collocate method (thin line) for steps 2, 3 and 4 of the map task. The initial steps (locating the number on the clock face) and the final steps did not differ (bars indicate one standard error).

The three dependent variables of path-length, fixations and duration were entered into a MANOVA, and multivariate tests showed that all factors and interactions produced significant effects (Figure 4). Because of the effects of Session factor, separate MANOVAs were then conducted upon each sessions' data, in a 2x5 repeated measures design. The effects of Method, Step and their interaction were significant in both of these analyses. Five further MANOVAs were then conducted upon each Step, pooled over both Sessions, with the single factor of Method. The multivariate tests showed no differences between the conditions for Step 1 ($F(3,18)=0.211$, $p=.887$) or Step 5 ($F(3,18)=1.16$, $p=.352$). Steps 2, 3 and 4 showed significant differences between the methods (Step 2 $F(3,18)=4.75$, $p=.013$; Step 3: $F(3,18)=10.9$, $p<.001$; Step 4 $F(3,18)=8.29$, $p=.001$), with the Translate condition resulting in longer path-lengths, more fixations, and longer durations for each of these three steps (Figure 4). Univariate tests within each of these MANOVAs showed the same pattern of effects for all three dependent variables.

Discussion

Despite the consistent placing of the clicked points at the centre of the screen following a transition to a new map, the translation interface is used less efficiently than the collocated interface, where the new point is in the same physical location as the old point, wherever this is on the screen. This is despite the fact that the place names themselves are not in exactly the same positions on successive maps, so that some search is required in both interfaces. The need to move gaze from the clicked location back to the centre is an impediment to the usability of the translation interface.

While watching participants use the maps, it became apparent that the typefaces used for the spurious place names did not look like the simple serif and sans serif faces used for real names on the original map (Table 1). This could have made them stand out more from the background, perhaps giving some advantage or disadvantage to one of the interfaces that might not be apparent in a real map. In fact, the differences between the typefaces might explain why the difference between the two interfaces is larger for Step 3 than for Steps 2 or 4. Step 2 is the first step on which the place name has to be found, and it was in a Helvetica Bold face. It then changes to Times Normal on Step 3, and on Step 4 to Georgia Normal. The appearance of the name is thus quite different between the second and third steps, but quite similar between steps 3 and 4. The change from Georgia Normal to Palatino Italic on step 5 is also quite different, although both are serif faces, and here the interfaces performed equivalently. We therefore ran the task again using different typefaces, changing between Times and Helvetica on each step, and including one change from normal to italic, one between two italic faces, and one from an italic to a normal face.

EXPERIMENT 3

Method

Twelve participants, all postgraduate students (8 males and 4 females; age range 21:5 to 33:11, median 28:5), took part in this experiment. They were paid £5 for their assistance. This experiment used the same materials and design as Experiment 2, except that the typefaces used for the spurious place names were altered to make them more like those of other names on the maps and the changes more systematic over different steps in the task (Table 1).

Results

Error rates in this experiment were very low, with only five trials needing to be repeated. Three of the errors were made by one participant, who also took abnormally long to complete several steps of the Translate task (over a minute in one case), and so his data were excluded. The three dependent measures of fixations, path-length and duration were computed for each of the five task steps in both conditions, as in the previous experiments. A MANOVA showed a main effect of condition $F(3,8)=10.8$, $p=.003$, η_p^2q=.802, Step $F(12,120)=8.1$, $p<.001$, $\eta_p^2=.448$, and their interaction $F(12,120)=2.28$, $p=.012$, $\eta_p^2=.186$. Separate MANOVAs at each step showed no effects of condition for Step 1 $F(3,8)=0.51$, Step 4 $F(3,8)=1.74$ or Step 5 $F(3,8)=0.58$, but significant effects for Step 2 $F(3,8)=10.8$, $p=.003$, $\eta_p^2=.802$ and Step 3 $F(3,8)=10.8$, $p=.003$, $\eta_p^2=.802$). One-tailed t tests showed that all three measures differed for Step 2 and Step 3.

Level	Experiment 2	Experiment 3
Step 2	**Helvetica Bold**	Times Normal
Step 3	Times Normal	*Helvetica Italic*
Step 4	Georgia Normal	*Times Italic*
Step 5	*Palatino Italic*	Helvetica Normal

Table 1: The typefaces used in experiment three were chosen so that they were more consistent than those in experiment two, were more like the faces used for real place names on the OS maps, and the differences between successive maps were equivalent.

Discussion

This experiment provides a close replication of the previous findings, despite the changes in typeface. The advantage of the collocated interface is the same on steps 2 and 3 (when the place name changes from a sans-serif normal to a serif italic face), but declines on step 4 to become non-significant (when the place name changes between two italic faces). In all three experiments, the initial and final steps do not differ. The initial step, of course, is in fact the same for both interfaces, because the place name has not yet been found and so there is nothing to collocate. On the final step in Experiment 1, the target was a symbol rather than the place name, and was not collocated with the point clicked and so the interfaces are again equivalent. However, in Experiments 2 and 3 the last four task steps all require participants to click on the place name, and so there is no obvious reason why the advantage of the collocated interface should disappear so consistently in the last step.

One possibility is that the low level of detail on these final maps makes it easy to re-orient on the place name, so that the searching required by the translated interface becomes trivial.

GENERAL DISCUSSION

The three experiments showed a consistent benefit for collocating the point of interaction with its result in a map interface. In summary, this is what May, Dean and Barnard [11] predicted, but which the designers of some online maps have not implemented. While it would be nice if the conventional design of maps were to change as a result of this work, it would also be good if it leads to a better understanding of how principles and theory can be used to guide design decisions in interface design. This research has followed a path from anecdotal observation (that film editors seem to control the position of elements within the frame) to a cognitive model of film watching (in which any demand for visual search interrupts the comprehension of narrative), to empirical data supporting the model (in terms of eye movements following cuts in a commercial film), and applied it to interface design (in the form of

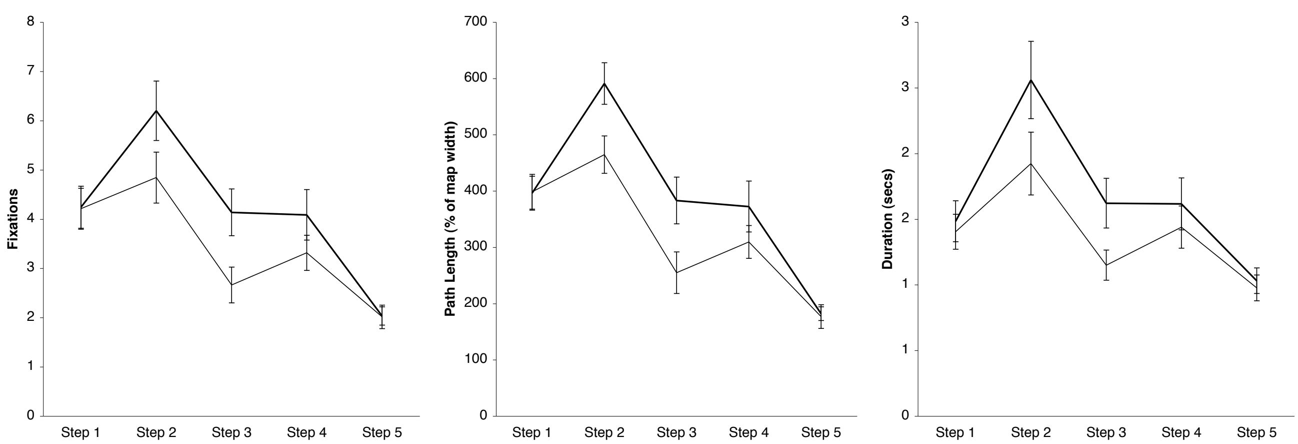

Figure 5: Changing the typefaces of the place names shown in experiment three did not remove the advantage of the collocated interface (thin line) over the translated (bold line) interface for the intermediate steps of the task (bars indicate one standard error).

contextualised but generic principles such as collocation). Here it has been shown that this principle does what the theoretical model says it will do, making it easier for a map user to understand what they are looking at when they click to zoom-in to a map. This paper is not so much an argument about the design of interfaces for maps, but about the utility of theoretically based principles for interface design. It also shows that the participants in this study were not benefitting from having learnt a consistent translation principle, despite its widespread adoption.

In the case of maps, it is not as if the designers are unaware of the confusion caused by the unfilmic, move-to-centre translation. Most online maps now try to guide the user's gaze to the new, central location by making use of a form of animation, interpolating four or five views of the initial map progressively moving a little each frame so that the clicked point moves toward the centre of the screen, before then displaying the new map. Whether intentionally or not, they are following the suggestion of Mackinlay, Card and Robertson [9] that the user should be supported in relocating points of view by moving them through virtual space rather than jumping directly from point to point; see Shanmugasundaram and Irani for an experimental study exploring this technique [15]. They are making their interfaces resemble fictional interfaces such as the 3D hologram viewer used by Deckard in Ridley Scott's film 'Blade Runner' [14]. This is not how film editors create their own material, however, because the experience of watching uncut footage shot while the cameraman is moving through a scene is not only time consuming but often nauseating, when the viewer's bodily and visual sensations do not correlate. Cutting directly between two different viewpoints can work, if collocation is used, as filmmakers do know, and as interface designers should know.

The collocation principle applies to more than map interfaces. It should also work whenever the user wants to move directly from one object to another, as yet unrepresented, object, and where the previous view is now superfluous. Clicking on an entry in a table of contents in a word-processor can allow you to jump to that heading in the body of the text, but the heading is rarely located on-screen in the same position as was the corresponding entry in the table of contents, and must be searched for.

Overall, this series of three experiments has shown that the advantage of collocating maps over the conventional approach of translating them to the centre is robust, supporting the argument that the design principle derived from the analysis of film cutting techniques can be used to improve interface design. Future work needs to show that the advantage of collocating is not generic, but depends upon the task being performed. There is some indication here that collocation does not always improve performance, when the search is simple in the final step, for example. May, Dean and Barnard [11] found that collocation was not used in film editing in situations where the location of the topic of interest was predictable and when the occurrence of the cut could be anticipated, as when two actors were talking to each other. In such conversational scenes, cuts would alternate between views of the two actors faces, on alternate sides of the screen. The resulting sequence resembled the view of the actors one would have if standing next to them while they were talking, moving ones gaze between them, prompted by the dialogue. Another class of cut in which collocation did not occur were 'over the shoulder' cuts, where an actor would be seen doing something before a cut, and the camera would then show the result of their action from behind them, with the rear of their head and a shoulder occupying a corner of the frame. Such framing of the action serves to provide a context, linking the results of the action to the person responsible, and also allowing one to see it from their point of view.

Extrapolating to computer interfaces, a similar lack of benefit from collocation should be found when context is important, or when translation of focal point after a screen change is predictable. Earlier in this paper, it was suggested that a scenario that might benefit from the retention of contextual information would be searching through a hierarchical database, where one is successively refining the category within which one is searching. Instead of replacing the initial set of categories with a second, finer-grained subset, a translated interface would place the new subset off to one side, with the original selection still visible so that the result of the interaction is contextualised by the super-ordinate category name. This form of interface is in fact used in Apple's iTunes Column Browser view, in which three panes arranged horizontally contain genre, artist and album titles. When nothing has been selected, each pane shows all of the information in the Library. Clicking upon a genre in the leftmost pane refines the content of the other two panes so that only artists and albums within that genre are now listed. Clicking again upon an artist further refines the album list to only those albums by that artist within that genre.

In a future study, databases could be contrasted that work in a similar manner to this, with the new information appearing alongside previous windows of information, with collocated designs where the new information replaces the previous window. If the collocation principle was working simply because it minimised eye movements, then the collocated versions of the databases should similarly benefit, because the new information is physically located close to the previous focus of attention. If, however, collocation worked in the Maps interfaces because it was concordant with the users' task, then it should not be as useful here, because the task involves successively moving through a data structure in a predictable manner, with detail unfolding on each operation. A translated design in which the new window opens to the right of the previous window will be contrasted with a collocated design in which the new window opens where the old window was, and the old window appears to the right. The collocated design thus removes the sense of moving successively through the data while retaining the availability of contextual information.

From a psychological point of view, the model of interface use and the ICS account of cognition can be used to reason about the cognitive consequences of changes to the task, such as the use of verbal material of increasing ambiguity (increasing demand upon a propositional-morphonolexical loop). Without the link between the principle and its theoretical derivation, such extrapolation and extension would be impossible and the guidance to designers would become rapidly out-dated, as new devices are developed, and novel tasks digitized. It also provides a link between the applied domain of human-computer interaction and computational models of visual search, (see, for example, Zelinsky [18]), where top-down models in which search is driven by object based information are simulating human behaviour with increasing accuracy.

ACKNOWLEDGMENTS
This work was supported by the UK Engineering and Physical Sciences Research Council grant EP/C515528/2.

REFERENCES

1. Barnard, P.J. Interacting cognitive subsystems: A psycholinguistic approach to short-term memory. In A. Ellis, (ed.), *Progress in the Psychology of Language, Vol. 2,* Erlbaum: London (1985), 197-258.
2. Bellotti, V., Blandford, A., Duke, D., MacLean, A., May, J. and Nigay, L. Interpersonal access control in computer-mediated communications: A systematic analysis of the design space. *Human Computer Interaction 11*, (1996), 357-432.
3. Burigat, S., Chittaro, L. & Gabrielli, S. Navigation techniques for small-screen devices: an evaluation on maps and web pages. *International Journal of Human-Computer Studies, 66,* (2008), 78-97.
4. Garlandini, S. & Fabricant, S.I. Evaluating the effectiveness and efficiency of visual varables for geographic information visualization. In K.S. Hornsby, C. Claramunt, M. Denis & G. Ligozat (Eds) Spatial Information Theory COSIT2009, *Lecture Notes in Computer Science 5756,* (2009), 195-211.
5. Harrower, M., & Sheesley, B. Designing better map interfaces: A framework for panning and zooming. *Transactions in GIS, 9(2)*, (2005),77-89.
6. Kraft, R.N. The role of cutting in the evaluation and retention of film. *Journal of Experimental Psychology: Learning Memory and Cognition, 12*(1), (1986), 155-163.
7. Kraft, R.N. Rules and strategies of visual narratives. *Perceptual and Motor Skills, 64*, (1987), 3-14.
8. Mackinlay, J.D., Card., S.K., & Robertson, G.G. Rapid controlled movement through a virtual 3d workspace, In *Proceedings of Siggraph'90*, (1990), 171-176.
9. Manson, S.M., Kne, L. Dyke, K.R., Shannon, J. & Eria, S. Using eye-tracking and mouse metrics to test usability of web mapping navigation. *Cartography and Geographic Information Science, 39(1),* (2012), 48-60.
10. May, J., & Barnard, P.J. Cinematography and interface design. In K. Nordby, P. Helmersen, D.J. Gilmore, & S. Arnesen (Eds.), *Human-computer interaction: Interact'95,* Chapman and Hall (1995), 26-31.
11. May, J., Dean, M.P. & Barnard, P.J. Using film cutting techniques in interface design. *Human-Computer Interaction, 18* (2003), 325-372.
12. Moellering, H. At the interface of cartography and computer graphics. *Proceedings of the 2nd annual conference on computer graphics and interactive techniques,* Bowling Green Ohio, June 25-27, ACM Press (1975), 256-259.
13. Monsell, S. Task switching. *Trends in Cognitive Sciences, 7,* (2003), 134-140.
14. Scott, R. (Director). Blade Runner [MotionPicture], United States: Warner Brothers (1982).
15. Shanmugasundaram, M. & Irani, P. The effect of animated transitions in zooming interfaces. *Proceedings of the working conference on Advanced visual interfaces '08*, ACM Press (2008), 396-399.
16. You, M., Chen, C.-W., Liu, H. & Lin, H. A Usability Evaluation of Web Map Zoom and Pan Functions. *International Journal of Design, 1(1)* (2007), 15-25.
17. Young, E., & Clanton, C. Film craft in user interface design. Tutorial presented at the *InterCHI'93* conference, Amsterdam (1993).
18. Zelinksy, G. J. Eye movements during target acquisition. *Psychological Review, 115(4)* (2008), 787-835.