Great Circles

HCI Smart Key Project

Project Phase 4:

Evaluation

Great Circles

TEAM
Keenan May
Amrutha Krishnan
Andy Pruett
Abhishek Nandakumar

CONTENTS

EVALUATION GOALS

Project Goals

Our overall goal was to facilitate car co-ownership by a small group living together using an ecosystem of car door interface, smart key, and online calendar. In the context of this evaluation phase, we raised both a set of social questions and a set of usability questions. We tried to address both of these in a limited amount of time.

Design of Study

Overall Experiment Goals

We wanted to test several elements of our prototype and their fitness to the problem space. These were foremost interaction in a novel setting in context, the intuition of what we perceived to be the most natural tasks in our problem situation, and the social implications of display and evaluating usage data, that is, testing whether the mental models that users constructed through interaction meshed with our own mental models as designers.

Study Goals

With this study we wanted to primarily focus on whether our interface worked, but also wanted to get indirectly at a few larger questions. Participants were asked to think aloud as they acted out individual usage scenarios, and evaluators asked follow up questions after each task. These questions assessed the ways in which the UI worked for each given scenario. They asked for details and confirmation about participant understanding and mental models ('what do you think would have happened had you done something else instead?') or hypothetical acceptance questions ('would you ever want to use this functionality, if it could mostly be automated?'). We tried to be attentive to the fact that our system involves more than just a key interacting directly with the screen, and incorporated tasks that were pure interpretation of system states, in addition to tasks involving moving in and out of the range of the device. Our goal was to test whether users were able to construct an accurate mental model of system functionality on the fly. As such, we tried to strike a balance between explaining enough to give participants the context necessary to frame questions (for example: "this light tells you something about the status of the car. What do you think it means?") without explicitly explaining how the system works. We wanted to be able to get a sense for how well our prototype fit the user's model of how a car-sharing system should work, in addition to getting feedback on how well it works as currently laid out.

We also wanted to know whether a direct mapping interface would be preferred over a 'scroll-and-select' interface, the idea being that one or the other could be easier and more familiar to use, depending on the user's approach to the tasks. Coming from a television remote control perspective, the arrows and enter mode could be more preferable. Coming from a car key fob perspective, the direct mapping commands analogous to the lock, unlock, alarm, and trunk release button set could be preferred.

To measure the social impact of our system, we created a set of charts based on simulated usage data and wanted to get attitude assessments of other users from the perspective of the participant dividing expenses using usage data. This would tell us whether or not collecting the data about time and distance allocation and sharing really had an impact on the feasible.

Task Measurement Method

We decided not to measure participant task times, for several reasons. The first was that all of our flows were relatively quick (none took over a minute, and many were almost instantaneous, involving no interaction with the car screen at all). What we were giving participants were scenarios rather than tasks, since we were also concerned with how they would approach a problem in the context of our system. We also realized that we could not both use a think-aloud protocol and measure task time, since the act of thinking aloud would artificially extend the time taken to carry out a task, and would not evenly effect every task and every participant. We also had made the decision to not support any flows other than the 'correct' one, which means that we could not measure task time in a realistic environment where errors were possible. That is, if a participant clicked the 'wrong' item, this would simply not work instead of letting them dive down into an incorrect subscreen; any times garnered from this procedure would thus have not been accurate reflections of novice system behavior. Finally, the fact that we were measuring novice behavior meant that task times might be meaningless, anyway, since this system would likely be used primarily by experts familiar with its functionality. This left us with a relative lack of quantitative data.

To gather quantitative measures for usability, we used Likert scale questions about the overall system in addition to a Subjective Usability Score assessment (SUS, CITATION) after completion of all tasks. In retrospect, it might have been a good idea to have participants rate the intuitiveness of each individual task immediately after performing it, as general SUS scores did not have the resolution for us to gain much insight from them in terms of what the problem points were. Overall, the SUS was a useful tool because it is a canonical, accepted, and neutral measurement of usability impressions. We planned to use it to assess within participant agreement about usability in this context.

Finally, we handed out a brief questionnaire in addition to the SUS. The goals of the questionnaire were to understand the how the users felt about the system in terms of ease of use, learnability, consistency and frequency of use. Additionally it was directed towards understanding social attitudes towards sharing the car. In order to understand this we even gave the participants a few graphs that depicted the car usage information and asked them to review the same

Task Description

We streamlined our list of tasks significantly from earlier experiment plans, and specified in detail how to describe them in order to give consistent input to participants. We also further generalized the scenarios made our original questions more neutral than originally planned.

Evaluation Task Scenario	UI Elements Tested	Measuring Instrument/ Question	Values to be Measured
You inspect the key in three different circumstances: away from the car, near the car, and within reach of the car door.	key indicator color scheme and icons	What do you think the color and presence of the indicators on the key mean now? Is it clear that the car is in communication with the key?	intuitive color and cymbol mapping, awareness of feedback and visibility
You are at the car exterior, knowing you are about to begin a scheduled reservation of the car	claim your trip' screen	is it easy to drive without doing a quick reserve?	think aloud, stops, errors
Use the key to select 'the top scheduled trip'	key/ UI mapping	is the control intuitive?, is it clear that you can scroll this far down?	think aloud, stops, errors
You have an errand that will require the car for about 45 minutes, you are at the car exterior	quick reserve sequence	How easy is it to quickly reserve the car + go?	think aloud, stops, errors
You arrive at your away-destination and exit the car	the 'exit/continue trip' screen	Is it clear whether you press end trip or not, mid-trip?	whether user does the right thing, think aloud
You are out on a trip, finishing the errand at your away-destination and you return to the car.	the 'exit/continue trip' screen	Is it clear that the car will auto-unlock here?	whether user does the right thing, think aloud
You are away from the car, you look at the key, and it shows some time formatted numerals. What does this mean?	the key time display	can the user interpret the time display?	Is user intuition interpreting this as time remaining or clock time?
You get home from your trip. What do you do?	the end/continue trip screen	is it clear how to end a trip/ whether it is necessary?	time taken, user commits proper action
you want to know what the car is doing later.	the display of public/ private events	Which events are joinable?	think aloud
You see that your roommate is going to the store later. Join their trip.	upcoming trips, join trip screen	whether its obvious how you would join a trip. Is there enough feedback to indicate you joined?	think aloud, stops, errors

Evaluation Task Scenario	UI Elements Tested	Measuring Instrument/ Question	Values to be Measured
You are on campus, your roommate(en experimenter) has the the car on campus also, and you want to get a ride home.	join in progress screen	how obvious it is that you can join. Does the user feel restricted by this screen?	think aloud, stops, errors
You are in the middle of a trip (on campus), and your roommate wants to catch a ride.	add person screen	whether it is clear how to add people/ is it clear who (on the display) is part of the trip, and who isnt	think aloud, stops, errors
you are in the middle of a trip, and you drop off your roommate (we role play the roommate). Get back in to drive without your friend.	the 'exit/continue trip' screen	is it clear that someone can just 'leave' a trip?	think aloud, stops, errors
At the end of the month, you want to divide up the costs of usage.	car info screen (paper simulation)	can a consensus be reached? Is there missing or extra info? Info easy to understand? , self-reported randomness of decision?	convergence of sentiment, and allocation amounts

Prototype Changes and Challenges

To conduct our evaluation, we had to allow our participants to experience all our test scenarios while using our smart key system. We extended the single all-inclusive prototype we designed to incorporate twelve scenarios. These scenarios were designed to specifically allow participants to think aloud. As a result, we were able to test all parts of the system - visible and invisible.

The process involved duplicating the initial trial application and including contextual information from each scenario that the design of our study described.

Apart fromt the two initial applications - Car Window UI and Smart Key, we now built a debugging controller for the car window, which would allow us to dynamically load the scenario that we wanted our user to experience next. We created fields to indicate the name of the participant and initial lock state. On submission of this form, the Car Window UI recognized these changes and reflected them immediately.

We had two variants of our prototype- each affording the corresponding input mechanism of the user either performing a conventional cursor movement on the interface using the smart key, or directly mapping the buttons on the smart key to the actions on the Window UI.

We created a lock switch in the debug controller that would allow us to change the lock status of the car door while transmitting information about the current context of the car to the smart key. This allowed us to simulate the car's awareness of the proximity of the user and the presence of the smart key. By illuminating the indicator lights on the smart key in various colors, we hoped to increase the visibility and transparency of the system's ambient behavior.

To allow for the accurate projection of the interface on the plane of the glass, we applied a 2D transformation using CSS3 that allowed us to tilt the application and confine it to the edges of the holographic film applied on the window.

Evaluation Setup

All aspects of the study were IRB approved. This actually forced us to think about our intentions and plan tasks and protocols in great detail, which helped us carry out the experiment.

As for subject recruitment procedures, we did not have any active exclusion policies for this study. We primarily recruited Georgia Tech graduate student from the Tech Square Research Building among friends and peers. In some cases we informally traded group members with other groups. While this is hardly an impartial sample, for our purposes it actually worked rather well, as some users voluntarily assumed the role of heuristic evaluator. However, these were generally people familiar with a variety of user interface conventions, and our results may have been skewed by this fact.

The apparatus for the study was our prototype, setup in a lab room in the Technology Square Research. It was our optically treated Nissan Maxima car door, a projector, two laptop computers, and an iPhone. We wanted to use an actual door in order to model the actual usage conditions. We also made sure that lighting was as strong and direct as as possible in order to simulate the effect of sunlight. However, could have done more to increase external validity; for example, we might have dirtied or fogged the surface of the door, used very strong lighting, or performed trials in exterior daylight. One laptop was used to serve the video to the projector and host the site. The other laptop was used to proceed through the preprogrammed scenarios in the backend and edit configuration files from behind the scenes.

Door - people understood that a door has certain innate interactions associated with the structure: opening, closing, operating the handle, etc. When participants were asked what they would do if they had reserved the car, they said that they might try (one participant actually reached for the handle) to physically open the car door to show what the real world interaction would be like. This helped us understand usage intricacies beyond what would have been possible had we limited ourselves to using paper and cardboard for the construction.

Line on floor. We wanted to be able to replicate how current smart key systems work. They already sense the driver's proximity to the car and help unlock the door when the user tries to pull the handle. To help users not familiar with smart keys understand how our system augments this experience, we decided to use yellow tape three feet away from the car door as a line to help simulate the range within which the smart key could be detected by the car. Ambient functionality such as having the door unlock or lock automatically based on the users' proximity were easier to understand by drawing this line.

Single iPhone: We prototyped our application so that only the driver of the car may be able to control the interface on their window. This allowed us to use a single iPhone to simulate all the interactions required for our trials.

Printed usage data: To lend focus on the web application component of our system that provides a calendar to reserve the car and see statistics on usage, we provided our users with the printed copies of the statistical information generated by the web application. We asked our participants to make choices of cost allocation and rate other users' behavioral aspects on a likert scale to understand how this information would be useful at a glance.

Evaluation Procedures

Recruitment: We used word of mouth to recruit participants, inviting them to try our project interface and telling them it was about a smart key.

Task Evaluation: Introduced the participants to the project concept and completed consent document.

Meanwhile, inserted participant name into the interface parameters to personalize the view. Presented the key/iphone to the participant and explained that it is a simulation. Presented the door and explained that it meant to symbolize a complete car in an outdoor or garaged setting. Pointed out that the yellow caution tape on the floor was not a warning, just an indicator of distance to the door.

Gave a description of the overall system and its context of use without explicitly describing the user interactions. We didn't say, "press button X". We set up the scenario with a short prompt, then observed and guided.

We then had them perform the various trials as discussed previously. While performing the tasks we asked the users to think aloud so as to know how they users perceive the system. At the same time we also recorded the think aloud so not to loose any information as well as to analyze it post the evaluations. All the participants were notified beforehand about the recordings and they all gave consent for the same.

Once all the trials were performed we gave each participant a set of graphs that depicted the car usage information and then asked them about their opinions regarding the same. In addition, we had users complete the SUS for the car door system, after the usage task was complete.

EFFECTIVENESS OF METHODS

Measurement Methods

While our evaluation generally served our purposes, there were significant pitfalls and perks to our approach.

As mentioned earlier, the majority of our data was highly qualitative in nature. We were not able to effectively code this data, so it became difficult to attempt objective evaluation of different aspects of the system. However, the focus on supporting open-ended response and feedback (due to the emphasis on questions and the decision to use think-alounds instead of timing tasks) gave us a wealth of useful design insights

Prototype Discussion

The key prototype brought a sense of realism to the experiment and users expected the action-reaction time to be immediate as one would expect with current smart keys. The time taken for the previous action to have its effect on the system resulted in an experience that felt less natural. Users had to repeatedly look at the prototype key to make sure they were pressing the right buttons when they felt their some of their actions did not have an effect. We understood that prototyping a physical key might have improved the experience of using the system.

While the SUS told us that users generally thought our system was usable and useful, this did not actually give us much useful insight, since, after all, our flows and tasks were extremely simple (often just one, or zero, clicks).

Some features of the prototype limited the effectiveness of our evaluation. Using a web based system to tie the interaction views together, we achieved high quality in the look and feel of the door UI, but low fidelity in the qualities that mattered for the key. In addition, the interaction design was skeletal and looked like it could do more than was actually functional.

The high fidelity of the look and feel of the door UI prototype we made captured the visual aesthetic, but the interactions were not completely programmed and did not allow completely free browsing and exploration. In our evaluation, we displayed UI views that implied many functions that were possible, even though these were not implemented. Indicators on the door and key were implemented, and feedback measured, but visibility of the indicators could be improved. The key prototype was relatively low in requisite fidelity when it came to evaluating the naturalness element of the key designs. In the script for our evaluation trials, we clearly state that the smartphone was a simulation of what should be a tangible key fob. However, simply saying so was no substitute for the real thing.

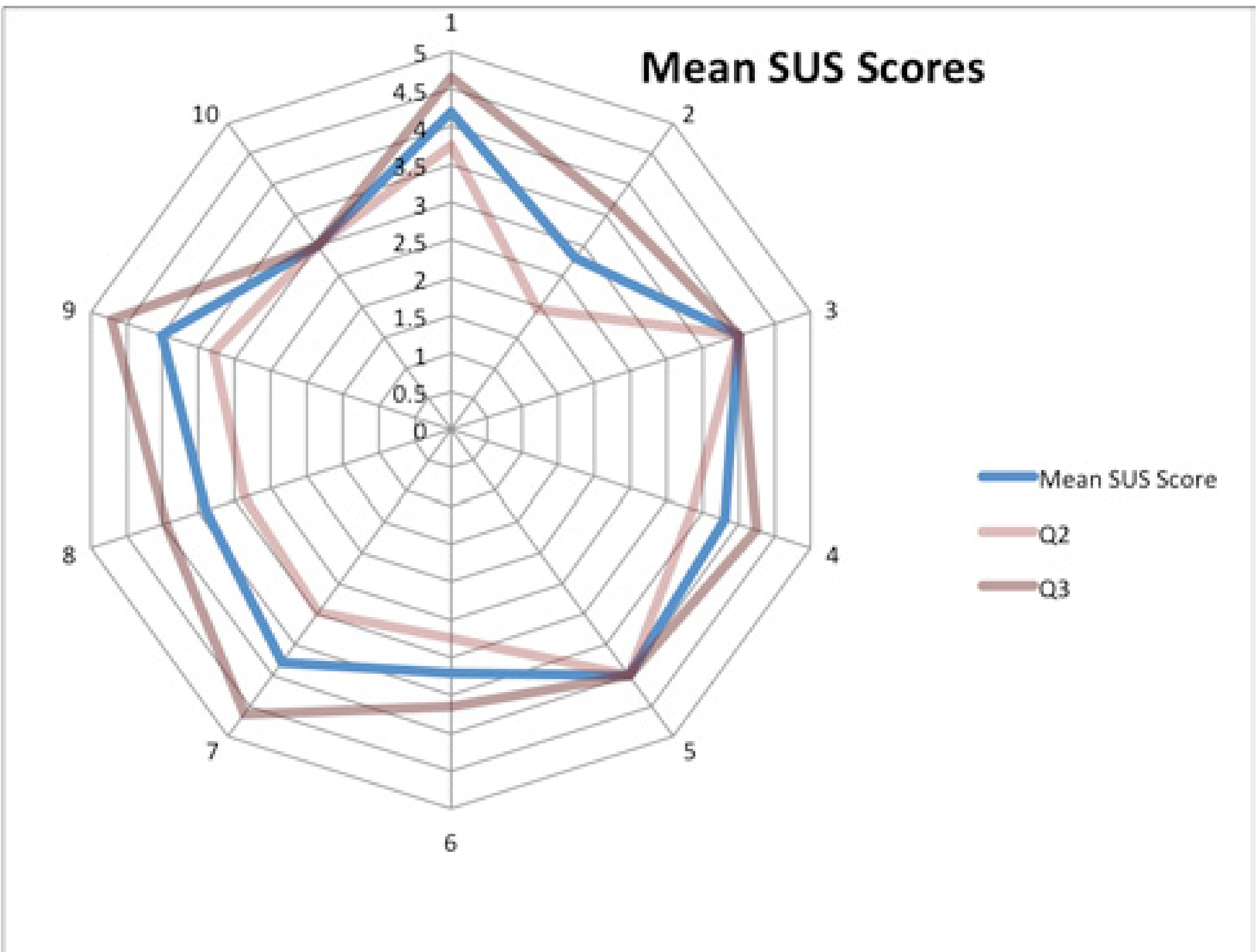
A real tangible object with pressable buttons and indicator lights would be a much better tool for the kinds of intuition and ease of use factors we were testing.

For example, many participants attempted to touch the status display icons on the key. We were unsure if this was due to the affordance of a touchscreen, the low contrast of the actual buttons, or whether these display items actually looked like they should have been buttons. One participant also experienced significant system lag, perhaps due to local wifi traffic, and had to have their session cut short. These issues represent external validity problems, but in our view did not seriously damage the quality of the results, as participants seemed tolerant of minor inconveniences, understanding that it was a prototype. However, evaluation of the alternate input method (absolute mapping) was hampered by the fact that any actual implementation of this method would employ tactile coding of button identity. The images of button shapes lacked the necessary feel quality. Several participants stated that this method would be more appealing and usable if buttons had a tangible shape so that they could be felt in low light conditions, from a user's pocket, etc. Thus, while our chosen prototype method did not allow us to correctly evaluate this interaction method, user feedback was still informative. Even the cursor movement design would have benefitted from tactile buttons. Touch screen displays in general require visual attention to correctly locate the active areas of the screen. When the participants had to divide their attention by looking back and forth alternately at the screen and key, their performance in the tasks was reduced and this reduced the evaluative strength of these tasks.

These all stem from a time and resources limit, and decisions about where to allocate those resources to fidelity along different dimensions of design. We felt that although there were gaps in our trial prototype, we were still correct in our decision to put effort into starting from a very extensible framework, a web UI, and fill in what we could along the way. Partly, we were still deciding what elements of the system would be most worth testing, and ultimately we only programmed the required tasks. Not only that, we budgeted time and effort to planning ahead in our design for inserting and deleting different interaction components such as sounds, external lighting, and sensor input. Working with an agile style in a very generalized prototyping framework enabled us to develop the minimum functionality necessary. For example, we had proposed using interactive scripting to recall system views and actions from a server. The goal was to allow a separation between experiment design parameters and the UI implementation. When time ran short, we were able to take the JSON data parameters from our experiment design and repurpose them as design documents for hard coded web pages. In this case, by designing for generality we were able to collapse our goals and adapt the working set of materials.

SUS Results

The SUS results were generally positive, with a high rating of around 4.0 with a standard deviation of within 1 point for each question. The result for question 2, "I found the system unnecessarily complex" was low (2.8) with a standard deviation of .84. In addition, question 6 "I thought there was too much inconsistency in this system" garnered a score of 3.2, with a standard deviation of .45. One potential explanation for these results is the inconsistency between tasks that require direct button input and tasks that require implied input (leaving/ entering the presence of the car), and the fact that some button interactions were perceived as unnecessary (add/remove, specifically). The highest rating was 4.2, for "I think that I would like to use this system frequently." This indicates that users can see themselves using the system, despite some inconsistencies.




60
hits
car
?
>
yes
free
is
available
until 1pm
to
know its you ? >
for
i would
sharers joining public ? >


29
hits
i would
think
be
have
probably
just
do
guess that
assume
say that
definitely

Usage Questions

Before completing tasks, users responded to some basic usage questions. Most users had not used a smart key, which meant that we could not leverage existing models of smart key functionality. Users were split on whether they currently owned a car, but all had driven at some point. No users had borrowed a car recently. While our sample group was too small to make strong statements about these user properties, it was clear from user responses that the concept of a proximity-sensing key was relatively unfamiliar.

Key Interpretation

During the first three tasks, which involved interpretation of the key states, participants had a fairly consistent set of problems. One general theme was that the meaning of the timer display was unclear- one participant interpreted this as the current time, and several others interpreted the timer as a specification of when the car would next change status. Several users did interpret this as a timer (which was our intention). It was also unclear that the color of the car light was linked to the status of the timer, and many of our users did not realize that we intended the timer to mean different things depending on the color of the car. In future design iterations, we will put more thought into the grouping of the car and timer, either by coloring the timer red, green or white or by visually grouping the car and timer in a more obvious manner. Users also tended to be unclear on the fact that the meaning of the timer changes when the car state changes. One user actually correctly interpreted everything correctly immediately. Our general impression was that the light states, if not the timer state, were actually fairly intuitive, but that we did not explicitly give the user the proper hooks to develop a good mental model. What this would entail would be adding a simple text description, or more descriptive iconography (perhaps a user 'in' the car or not, and the car in a 'garage' or not, harkening back to our previous phone app design). One user did report wanting a simpler scheme of green and red car states, which is something we might consider. However, this user was able to ascertain the meaning of the states with some exposure. This user could not see the purpose of having three car light states, specifying that her task was to know whether she could go get in the car and drive at a given moment, or not (regardless of reservation). It is apparent that information about the real status of the car 'at home' or 'not at home' should be separate from information about reservations and intent. In addition, we found that users were unable to interpret the meaning of the key light, either failing to notice it entirely or thinking that it was an unlock button, or other indication of unlocking privileges. The use of the 'key' visual tied this too strongly to notions of locking and unlocking, and provided too strong of an affordance for something that was meant to be a passive information display. As such, in subsequent iteration of this system, we would try separating out the actual car status indicator from the indication of usage intent. We would have one light that used the explicit, descriptive iconography of a car in a garage or not in a garage (a bit hard to do with backlighting and no display, but possible) to indicate the presence of the car. We would eliminate the key light entirely, since the connectivity is implied by the fact that the screen turns on (we might add some wireless bars at the end of the remote in order to keep this clarifying element, but it is not necessary based off of user feedback). The second light and timer would be spatially grouped, and would work in the following way. There are three backlit text areas reading 'next', 'left', and 'free until'. In concurrence with the expectations of most of the users, the time display would give an absolute time, and not a relative one. However, when a user's reservation begins, the time display would switch to a countdown, and begin strobing. This gives a clear indication that time is slipping away, and that the user should be taking advantage of their reservation. This was the whole point in showing the 'green' car state in the first place, but this was not clear to our users. Our main takeaway from the trials testing key state interpretation was that while the mental model we were asking users to construct was an intuitive one, we did it in a way that was too symbolic and inferential as opposed to descriptive, and tried to have simple iconography carry to much subtle information content. The new schema would support the variety of user approaches to a greater extent, and provides less room for differing interpretations.

Basic Interactions

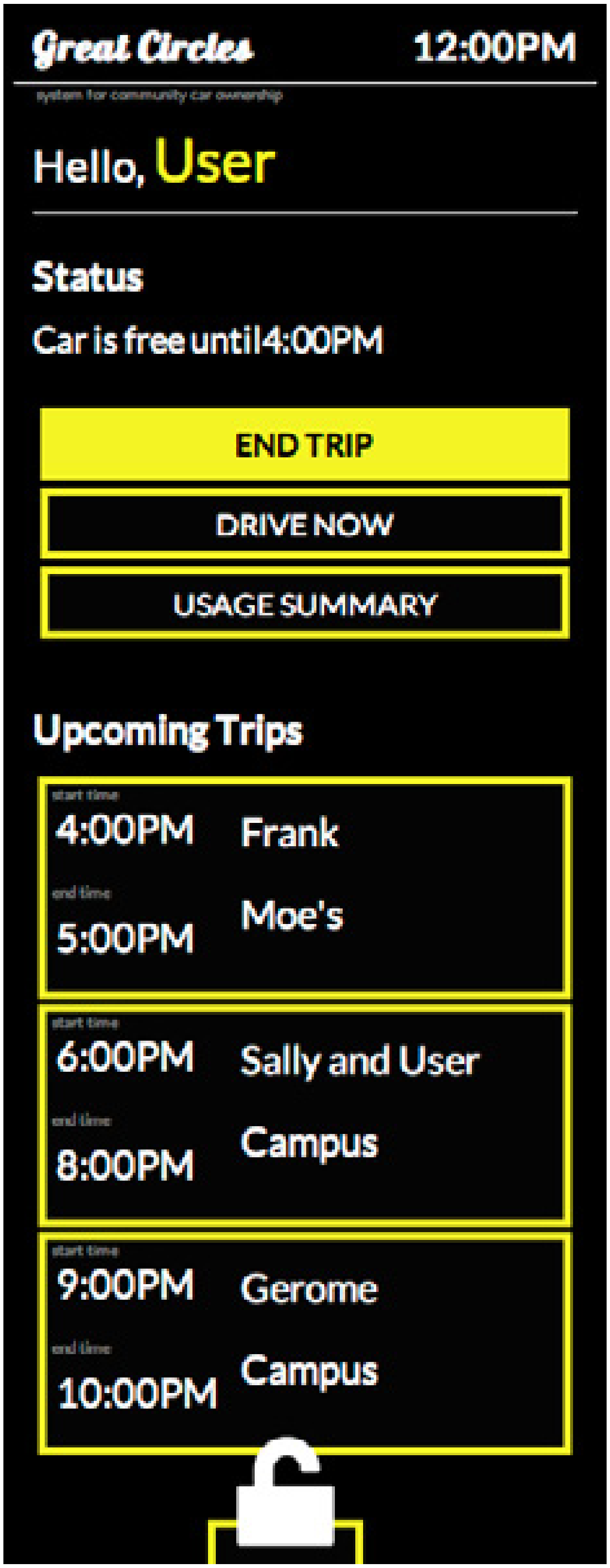
The fourth task, in which users simply had to approach the car and discuss their reaction, revealed (as mentioned earlier) that the status of the key connect icon was ambiguous. One user did not even notice the change, even after several steps forward and away from the interface. Some users were aware of the meaning of the key, but we realized that its information content is redundant, since the window turning on indicates a good connection anyway. Users were generally clear that the window had 'sensed' their presence.

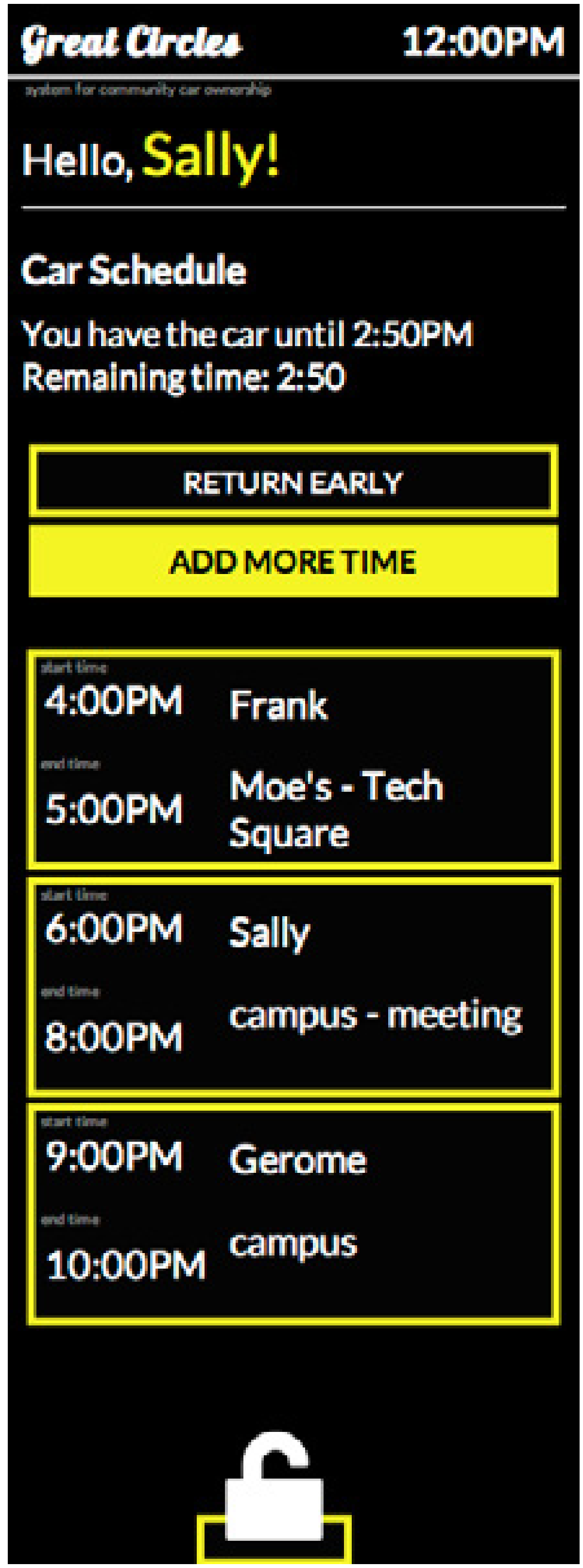
The fifth task was simple selection with the scrolling key. Users were quick to master this interaction style; however, two issues did arise. First, one users expected there to be 'wraparound' selection, and was surprised to find this absent. Another user was unclear that the 'square' button on the key was a for selection; as such, our key redesign mockup changes this button to explicitly read 'select' and to be shaped less like the 'stop' function on a media player.

'Drive Now'

Task six, which was to 'drive now' when a reservation had previously been made online, brought up several interesting points. Users were generally very comfortable with the flow, which is literally one or two clicks. However, several users did not see the purpose of 'drive now' in this context, and specified that they would simply enter the car if it had been reserved for them. This brought up an interesting point, which was that 'drive now' essentially had no purpose in this context, other than to force users to spend a moment attending to the interface and potentially notice schedule details. Most users did appreciate the presence of the 'got it' page in that it added confirmation and feedback, however. This led us to ways of considering how to eliminate 'drive now' when it was not appropriate while still providing adequate indication of how long a reservation was for and when it would end. Finally, some users wanted more feedback that a successful click had taken place, rather than an instant page change.

Task seven, in which a user reserves the car for 45 minutes, generally had a flow that users found very simple. However, some users went to the calendar interface first, intending to add an event either in or directly before or after the calendar events This led us to the realization that we needed to support more flexible joining behaviors at the door. One user wanted to be able to quickly reserve the car from now until the next reservation, or to reserve all of the time between two reservations. While it is unclear whether this would support the kinds of conscientious social behaviors we want to encourage, supporting this flow would definitely be worth testing. Essentially, there would be a '+' function between each calendar reservation (and before the first one)


Great Circles
12:00PM
system for community car ownership
Hello, User
Status
Car is free until 4:00PM
END TRIP
DRIVE NOW
USAGE SUMMARY
Upcoming Trips
start time
4:00PM
Frank
end time
5:00PM
Moe's
start time
6:00PM
Sally and User
end time
8:00PM
Campus
start time
9:00PM
Gerome
end time
10:00PM
Campus


Great Circles
12:00PM
system for community car ownership
Hello, Sally!
Car Schedule
You have the car until 2:50PM
Remaining time: 2:50
RETURN EARLY
ADD MORE TIME
start time
4:00PM
Frank
end time
5:00PM
Moe's - Tech Square
start time
6:00PM
Sally
end time
8:00PM
campus - meeting
start time
9:00PM
Gerome
end time
10:00PM
campus

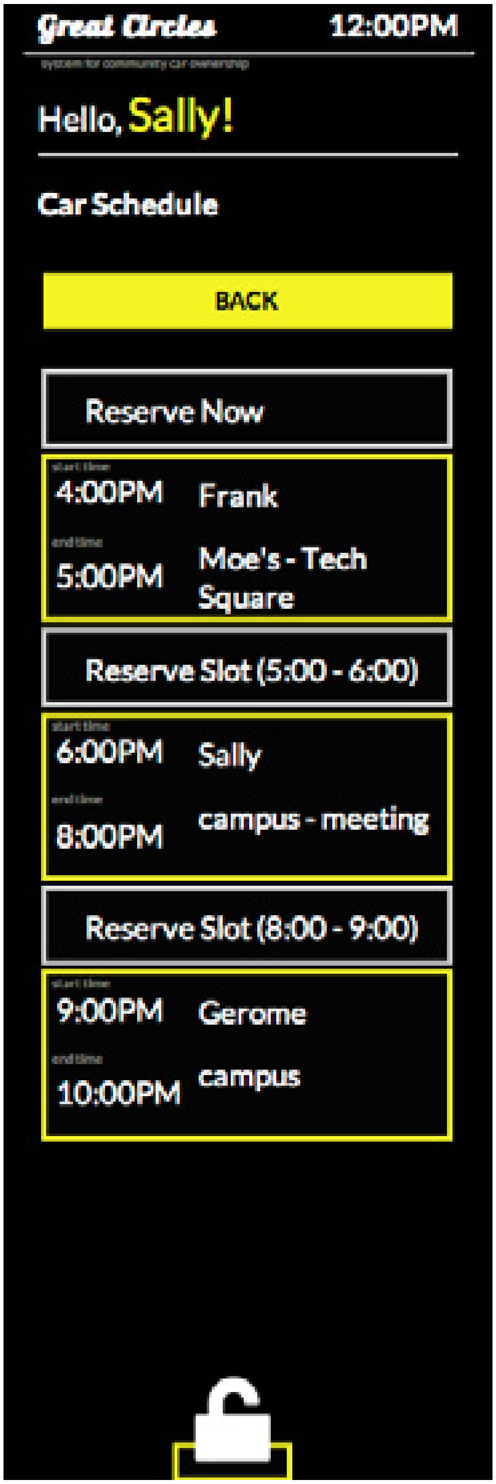
that would allow a user to quickly reserve the car during a certain future time slot. This is slightly more functionality than we wanted to support on the screen, but it is not difficult to do.

The question of when and whether participants would want to use the window interface garnered some very useful and varied responses. Many said that they would schedule important events ahead of time using the calendar, but that they would appreciate being able to simply get in and drive. However, it was not entirely clear to some participants why they would want to reserve a car for 'now,' instead of just driving. One participant reported that they simply disliked having to use online calendars, and wanted to avoid them wherever possible. Some participants expected a physical lockout if it was not reserved (even given the lock status icon being unlocked). As such, for subsequent designs, we will change the text of drive now to something like "estimate trip length." 'Drive Now' does correspond to the user's task, but, as several users mentioned, if they truly wanted to 'drive now,' they would just get in the car and go. Thus, we need to distinguish the 'drive now' functionality (which, after all, was conceived during a period in which there were hard lockouts for users who had not reserved the car) from simply getting in and going. We want to do this by highlighting the purpose of 'drive now,' or of scheduling trips at all: to communicate your intentions to the rest of the group. In our original designs, the interface was used primarily to gain access to the car, or permission to use it. During the prototype evaluation we realized that, by settling on schema in which the car was always unlocked, we had invalidated some of our earlier design choices. Thus, based on a lack of clarity (as reported by users) as to why they would actually want to use some of the functions (specifically add/remove members, and 'drive now' if a reservation already exists), we decided that, in future iterations, we would make it clear that the only reason you need to use the interface is for time estimation and to let other users know you want to come along on their trip, so that they will not leave without you. It's purely about being clear and courteous about your intentions, not about getting past restrictions to use the car. As such, we renamed 'drive now' to 'add drive time,' since this is more descriptive of the actual reason for clicking this. In addition, while the car is reserved for a user, that user will see 'return early' and 'add time' buttons. The purpose of both of these is to clarify to other users that they will be either early or late. This supports the existing behavior of texting someone to let them know you have arrived earlier than expected, or will arrive later than expected. It also makes sense in the context of the redesigned key display, since users will be able to see the car's state change from reserved to not reserved, regardless of whether the car is actually 'at home' or not.

‘End Trip’

The issues raised in subsequent tasks were similar to those raised by task seven. Task eight, which was to leave the car without ending a scheduled trip, was interpreted correctly about half of the time. Users generally wanted to simply exit the car, but were unsure whether ‘end trip’ would cancel their reservation or not. Once again, this was a case where our general interaction model was in line with the expectations of several users, but our labeling of functions was not descriptive enough to communicate how the system would actually work. While most users figured out how this worked over time, it should have been apparent from the start. Several users did not infer that stepping away from the car would take care of anything for them. However, after being told of this functionality, all users reported that they would want the system to handle locking automatically, and that they would be comfortable with an automatic locking system given solid feedback. Feedback would have needed to be in excess of the lock icon changing and the screen going off; users wanted something like a sound and/or a flashing of car lights in order to be extra clear. However, it’s important to note that users trusted the system to perform its function in this case. In future iterations, we would likely add functionality where the key vibrates twice to indicate the car has been locked, and once when the car is in range to indicate that the car has unlocked. Users were clear that they wanted better feedback on whether the lock/unlock had occurred or not, and the obvious issue of indicating a lock when a user is looking and walking away from the window (which came up during testing) is addressed by having a notification on the key itself.

Task nine was fairly well understood, probably due to the fact that it followed task eight, and thus users already had an accurate working model of the nature of the leave car/end trip distinction. Some users had trouble with the notion of ‘end trip,’ wanting a more representative label (‘end reservation’ or ‘give back car’). Users generally understood that a ‘trip’ was a reservation. However, it became clear to us that we need to strongly distinguish the reservation-centered interactions with the car from the facts of whether a user was actually driving it or not. The language of ‘end trip’ ties the act of driving to the act of having the car reserved, which is not a real functional division in our system. In order to make this more clear, as mentioned earlier we will be removing ‘end trip’ and replacing it with ‘add time’ and ‘return early.’ These labels both make reference to time (important in communicating why the screen interactions are actually meaningful over the automated ones), and clearly delineate what is actually going on with the system. This would be subject to further testing of course, but this is the direction our response to feedback on this point will go.



Calendar Function

Task ten was to join a fellow car-sharer's future event. Of note was the fact that many users did not know that the calendar events were clickable until it was suggested that they click them. In our next iteration, we will remove the gap (and, in fact, the functional distinction) between calendar events and other options such that creating a reservation is seen as creating a calendar event. All users had noticed the calendar, which indicates that our goal of increasing the visibility of upcoming events was well served by having the calendar visible. All users indicated that they would be okay with users joining their scheduled trips, if the option existed to make them private. However, several users mentioned that they would want to have the person whose trip had been joined receive some sort of reminder when they were about to leave so that they did not leave the 'joiner' behind. We plan to incorporate this into further iterations. One user also indicated that he was seeing the upcoming trips area as an actual calendar, and wanted to be able to quickly schedule a trip between any of the existing trips. While this did not come up with any other user, part of this may be due to the fact that we never tasked users with scheduling a time in advance. We will incorporate this feature into the next prototype iteration.

Adding and Removing

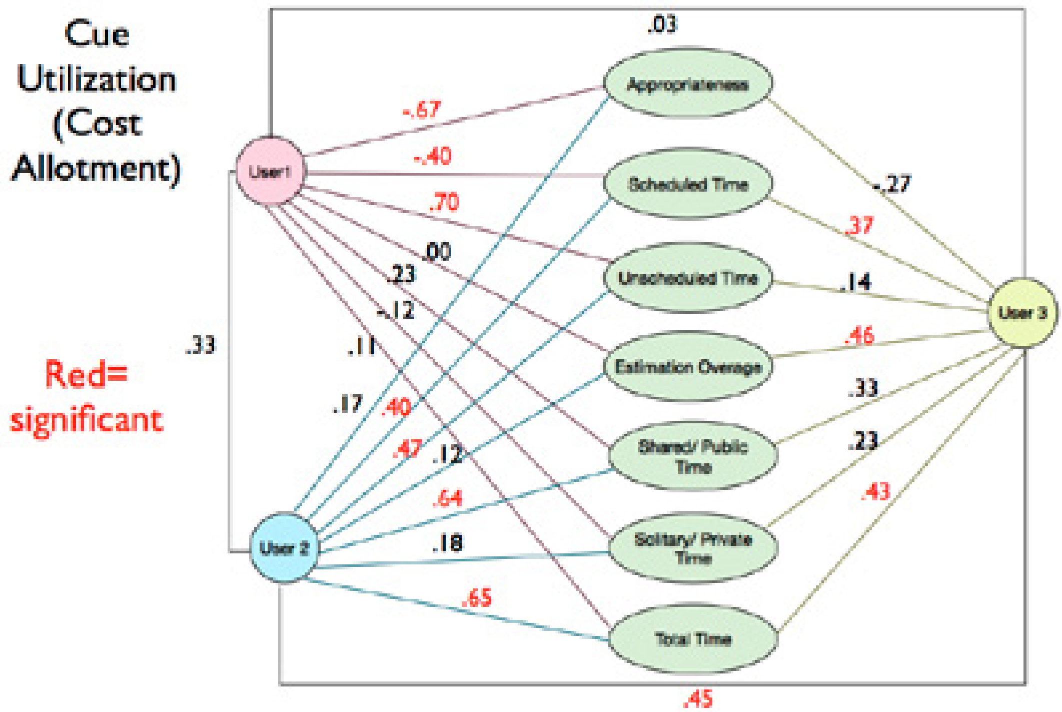
Tasks eleven and twelve, in which users were tasked with adding a friend and removing themselves from a trip, received the most negative feedback, overall. Almost every participant thought that these actions should be handled automatically by the system, and several voiced confusion about who would 'inherit' a reservation if its original owner left. Since adding or removing a person does not actually give the system any useful information (all usage data is tracked simply by who is present, and no time information is communicated here), we decided that it should simply be removed. This makes a few assumptions about system intelligence, but we decided that this needed to be a system logic problem. The usability problems can be solved by automating add/remove features and adding clear feedback (which is what users expected in the first place).

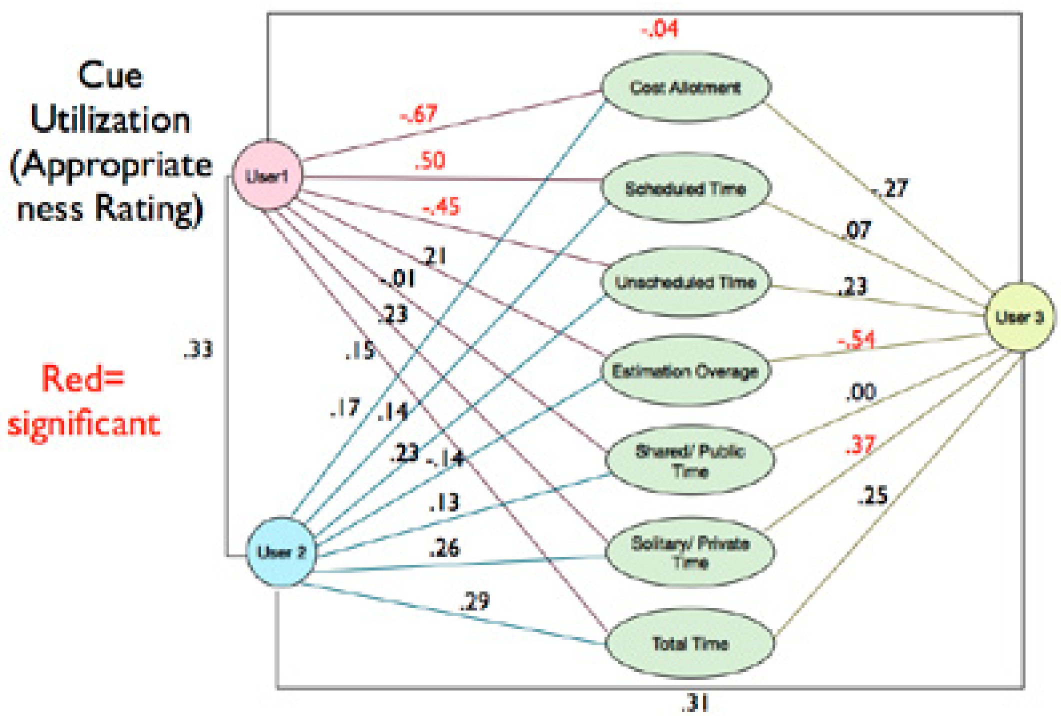
Absolute Mapping

Tasks thirteen and fourteen involved participants using the absolute-mapping version of the key. Participants tended to have trouble with this version at first, "It's a little confusing because of the previous app had the same controls… if I had not done that earlier it might have been better, but now that I had done that it was just confusing." During future trials we would definitely vary the order in which the two control methods were presented. Around half of the users clearly preferred the absolute mapping method. Several users commented on the fact that they would have preferred the absolute mapping method if they had used a key with tactile feedback. One user reported that the shapes chosen were too arbitrary, and that the arrows had confusing connotations. Future versions of this key style would include something tactile and logically ordered. Most users mentioned that it was awkward having to look back and forth between the key and the screen, but a large subset of these stated that they would probably become experts overtime, at which point they projected that they would prefer the absolute mapping method. All in all, it is unclear which method was better. In the future, we would like to develop the absolute mapping version further, and get user feedback after many trials of use (once the mapping had been memorized). One of our goals was making the system unobtrusive enough that it would be used, and the absolute mapping key version might serve that goal well; however, it was difficult to test this in the context of our fifteen minute study, as its true benefits, (if they exist) would have come as users reached expert level.

Usage Data: Methods

Properly evaluating the manner in which car sharers utilize usage data to make judgments proved a more difficult task than we first assumed. In further iterations of this system, we would want to find out whether each group might develop their own social norms. This was why our original intent was to run a focus group discussion session around each set of usage data- a side hypothesis being that each group would converge on a certain cue utilization scheme, thus validating our decision not to automatically produce a cost allocation suggestion based on some sort of algorithm. If the weights attached to each value indeed differed by group, no single algorithm could properly predict manual cost allocation. In addition, if some information was completely unutilized, it might not be meaningful to display at all. We attempted to evaluate the point of cue utilization during the usage data interpretation phase of our experiment. Early results indicated that cost allocation depended primarily on time used and distance traveled, as opposed to sharing ration and openness ratio. However, it became apparent that our initial method of having each participant allocate usage costs to four fictional car-sharers would not give us the kind of data needed. We hoped to be able to apply Brunswick's (1952) lens model in order to measure predictive power between each of the usage datum and judgments of cost and social appropriateness. We realized that we would have to simplify the judgments of social factors into a single factor, called 'appropriateness,' and that we could have participants rate the fictional car users in terms of both this and expense accountability. One thing we wanted to ascertain was whether social and expense accountability were highly correlated or not. One of our system goals was to support both methods of social norm enforcement. It was apparent from our initial results that usage data was not getting utilized to a large extent- the largest predictors of cost appeared to be simple usage time and distance. However, this was based off of only one set of usage statistics, so is difficult to make any strong claims about this. In order to find out with greater certainty whether this was a valid result, or was due either to the graphs being too complex or to the fact that raters only saw a single graph, we chose to retool and split up the graphs in order to facilitate distinct evaluation of each cue. We then redistributed a second survey that would give the number of responses necessary to form the necessarily cue correlation values to do a multiple-judge lens mode linear judgment analysis. This survey involved a small number of participants (three). Usage data was streamlined in the following fashion: Graph one displays scheduled and unscheduled time, per user, graph two displays usage estimation accuracy (lateness or earliness), and graph three displays shared/ public scheduled usage and private/ solitary unscheduled usage, per person. We made the decision not to include any display of distance traveled for this iteration of the study, since the goal was to provide simple and interpretable information. Raters evaluated data for a total of ten different data sets, and both allocated costs and answered a question about 'social acceptability.'


Cue Utilization (Cost Allotment)
Red= significant
User1
User 2
User 3
Appropriateness
Scheduled Time
Unscheduled Time
Estimation Overage
Shared/ Public Time
Solitary/ Private Time
Total Time
.03
-.67
-.40
.70
.00
.23
-.12
.11
.33
.17
.40
.47
.12
.64
.18
.65
-.27
.37
.14
.46
.33
.23
.43
.45


Cue Utilization (Appropriateness Rating)
Red= significant
User1
User 2
User 3
Cost Allotment
Scheduled Time
Unscheduled Time
Estimation Overage
Shared/ Public Time
Solitary/ Private Time
Total Time
-.04
-.67
.50
-.45
.21
-.01
.23
.15
.33
.17
.14
.23
-.14
.13
.26
.29
-.27
.07
.23
-.54
.00
.37
.25
.31

Usage Data: Results

After getting the results of this second usage data viewing session, two linear judgment models were constructed in order to investigate cue utilization: one for cost allocation, and one for ratings of social acceptability. There were a total of seven cues evaluated: total hours spent, hours spent in private/solitary usage, hours spent in public/shared usage, hours spent in scheduled use, hours spent in unscheduled use, and hours over/ under estimated time. Correlations were calculated between each cue and final judgment amounts, and cue utilization was calculated for each of the three users. Cues that were statistically significant ($p<.05$) were judged to have been utilized. We also calculated correlation between social acceptability ratings and cost allocation ratings. Finally, policy agreement was calculated between the three users. It should be noted that there is no 'ecology' side to this model, since there was no way of gathering data on what was 'correct;' the goal was simple to compare the factors the judges considers when making cost allocation and social appropriateness decisions using our displayed data.

In terms of cost allocation, User 1 attended mainly to scheduled time (r = -.40) and unscheduled time r= .70). They attached very little significance to overage,and minor significance to shared/public time and solitary/ private time. Interestingly, total time was not a very good predictor of cost allocation in and of itself. In addition, scheduled time was actually a predictor of reduced cost allocation, indicating that this considered the relative ratio of scheduled to unscheduled time rather than the raw magnitude. This is also reflected by the fact that, for this user, total time was not a good predictor. User 1 utilized cues in a similar manner to gauge social appropriateness: scheduledtime (r=.50) and unscheduled time (r=-.45). Cost allotment and social appropriateness were inversely related, and strongly correlated(r=-.67).

User 2 utilized different cues in both cases. They relied on scheduled time (r=.40), unscheduled time (r=.47), shared/public time (r=.64) and total time (r=.65). User 2 did not appear to have a discrete strategy for judging social appropriateness, as none of the cues were significant predictors.

Finally, user 3 utilized the cues of scheduled time (r=.37), overage (r=.46) and total time (r=.43) in order to determine cost allocation. User 3 looked at overage (r=-.54) and solitary/private time (r=.37) in order to determine social appropriateness. Social appropriateness did was not well predicted by cost allocation for this user.

Rater agreement was high and significant between user 2 and user 3 for cost allocation. There was actually significant disagreement (r=-.40) between user 1 and user 3 for social appropriateness, perhaps due to the fact that user 3 utilized entirely different cues from user 1.

What we learned from the lens model analysis was that our displayed data generally allowed users to make informed decisions about cost allocation and social appropriateness. However, each user utilized different cues, implying that developing a cost allocation would not work well unless its parameters could be tweaked to allow for different social norms. The only cue attended to by all three users was scheduled time, indicating that our efforts to maintain a scheduled/ unscheduled distinction was meaningful to this user group. Further testing would have to be done to determine if this was true at a larger scale. Users were most varied in how they considered overage- users 1 and 2 barely considered it at all, while user 3 attached high value to it. This indicates that either our display method was confusing or there is a genuine difference in how users value this characteristic. While this experiment considered individual rater behaviors, no attempt was made to simulate group judgment, as had been our original intent. A full analysis of such habits would be a study onto itself, but what we have indicates that tracking and displaying car usage, but not generating cost recommendations, is an approach that allows users to adopt a variety of value systems and judgment strategies.

FUTURE DIRECTIONS

Our questionnaire revealed generally positive attitudes toward the system. Importantly, users replied with the likert rating of 4.2 (5 being 'strongly agree') when asked if using the system would make them more likely to share the car. We received additional feedback at the two poster sessions, informing broad directions we could take this project. The poster sessions were helpful in terms of getting some really good feedback from the reviewers. A lot of them could actually relate to this problem and felt that this system would enable them to share the car effectively. An interesting point that came across during the poster session was the need to actually have a smartphone used as an alternative to the smart key fob. Though one of our primary concern with designing the smart key fob was to give the user a sense of ownership some of the reviewers also felt in the future the whole concept of a key would get replaced a single smart device like a smart phone. Though these suggestions might have been also due to the fact that we had used a smart phone to simulate the smart key fob these feedbacks have also prompted us to consider using the smart phone as an alternate way to interact with the window interface.

One suggestion that we heard repeatedly was the desire for an intelligent trip aggregation suggestion service. This had been considered in some of our earlier prototypes, but we chose not to go down this route, largely because in the context of a single car with a few users, it may not be particularly useful. If we were to generalize this system to larger groups with several cars, this sort of intelligent suggest system would definitely be worth incorporating into further design iterations.

After refining the usability of our interaction paradigm and data display with the results of the present evaluation, moving forward we would want to develop a higher fidelity prototype of the system and deploy it in a small community of car sharers.

REFERENCES

Brunswik, E. (1952). The conceptual framework of psychology. Chicago: University of Chicago Press.