User Testing Analysis for BlueSky

We found that the vocabulary we were using affected users' experience and understanding of the application. Users had trouble understanding what the 'description' of the intention should be and found 'intention' itself to be a bit vague. Once we debriefed the users, they were better able to understand the goal of our application so the 'intention' of our application will be something to focus on further as we develop. Overall, users found the questions to be insightful and relevant in helping them become more aware of what they were actually trying to get done. As the functionality is fairly straightforward, users found the flow to be intuitive and made sense to our users.

List of changes

- Change/remove name of the 'duration' label (for A/B)
- Change the name of the 'send' button to 'submit'
- On the home page or the 'Start your intentions page' add a little blurb about what 'intention' is and provide information about how the application is intended to help them.
- Change the name of 'description'
- Order the intentions by chronological order
- Remove the friends page
- Make stats page functional
- Switch 'task' with 'intention' in remaining instances

Patterns among users

- Many of our users were not sure by what we meant by 'intention'. We noticed that a common trend amongst our users was that they were confused by what we meant by an intention. Some of them were even confused by why we were asking them to create an intention in the first place. We think the users faced this problem because our homepage does not describe what an 'intention' is and does not describe what the main purpose of our application is. The users understood that the homepage was where they created an intention, but many of them did not understand why they needed to create an intention and how creating an intention could benefit them.
- All of our users were confused by what the 'send' button did after they answered the questions. Some of them were worried that their answers were being sent to another user on the 'Friends' page. We have decided to remove the Friends page completely from our application since it is no longer relevant to our core functionality.
- Most of our users were confused by what we meant by 'description' and 'duration' on the 'Create Intention' page. We believe the users had this problem since our terminology was vague and slightly confusing.
- Our users believed the questions page was an interesting feature and helped them hone their thoughts into the intention that they were describing. Our users reported the questions as "deep" and "insightful" in helping them reflect on their intention.

Fixed bugs and added features

- Added an intermediate page after the user submits questions for the first time for a given intention, enabling them to reflect further by asking additional questions.

- Changed "task" in the questions to "intention"to maintain coherence in terminology

Redesigning 'Start your Intention' Page

We will be redesigning our 'Start your Intention' page. Instead of simply having a label which says 'Intention' we will be asking the user the question 'What is your intention'. Also, in our redesign, we are going to remove the 'Duration' and 'Description' and in its place, we will just have one label called 'Elaborate more.'

All of the users we tested this week were confused by most of the terminology on the 'Start Your Intentions' page. They were confused by what we meant by 'description' and 'duration'. Since that page is crucial to our core functionality, we decided to redesign it and improve the terminology that is used on that page and make it more concrete. In Version A (our original version) the users did not understand the point of having 'description'. Currently, the terminology is inhibiting the users from effectively using the application so changes are essential if we wish for the users to be able to use and benefit from our application.

Version A vs Version B

These pages are accessible from https://on-purpose.herokuapp.com/new-task_A.html and https://on-purpose.herokuapp.com/new-task_B.html

Version A: Version A allows the user to input 'intention', 'description', and 'duration' upon 'starting an intention'. This works to break up each important feature of getting started on an intention and allows the user to see and input the relevant information.

Version B: A problem brought up in our testing had to do with the terminology that we used within the app. The initial 'start an intention page' will now bring up a text field with 'intention' and the option to 'elaborate' upon that intention. This method combines 'intention' and 'description' to streamline the process and has done away with 'duration' altogether. We determined that duration wasn't needed and the intention/description could be streamlined further.

We will measure the effect of these changes by recording the amount of text entered in each field. The principle underlying this is that a greater quantity of text correlates with more time spent reflecting on the intention. For A, we will look at the total number of words in both 'duration' and 'description', while for B we will look at the number of words in 'elaborate.' While in both cases it is possible for users to enter nothing, as the only check is that 'save' has been pressed in each field, we do not consider this a likely option, as none of the users we tested for this phase thought to leave them blank.

Hypothesis

- Version A is the original design and displays 3 important features that we, as a team, believe to be important in starting an intention. The 'intention' itself, a quick 'description', and the 'duration' of the intention itself. By displaying these 3 features promptly to the user, it helps them bring their thoughts solely to the intention at hand.

Despite the redesign of this page as version B, it's possible that that version is not better because there are other unaddressed issues with the design of the page or inherent difficulties in the task itself.

- Version B streamlines the 'start an intention' page and reduces ambiguity within the page itself. We have removed 'duration' (a feature found to not be as important) and have changed the 'intention'/'description' feature. Now, we have two textfields: one to define the intention and another to elaborate and describe the intention further (which can encompass any pertinent information that the user needs for that particular intention).

Compile & Analyze

We conducted an experiment using Google Analytics. Over two days we had 31 unique users in 38 sessions. This is shown in the first image below. The second image shows the progression of events by users from each page. The total displayed there likely differs from that in the first image because some visitors did not trigger any events.
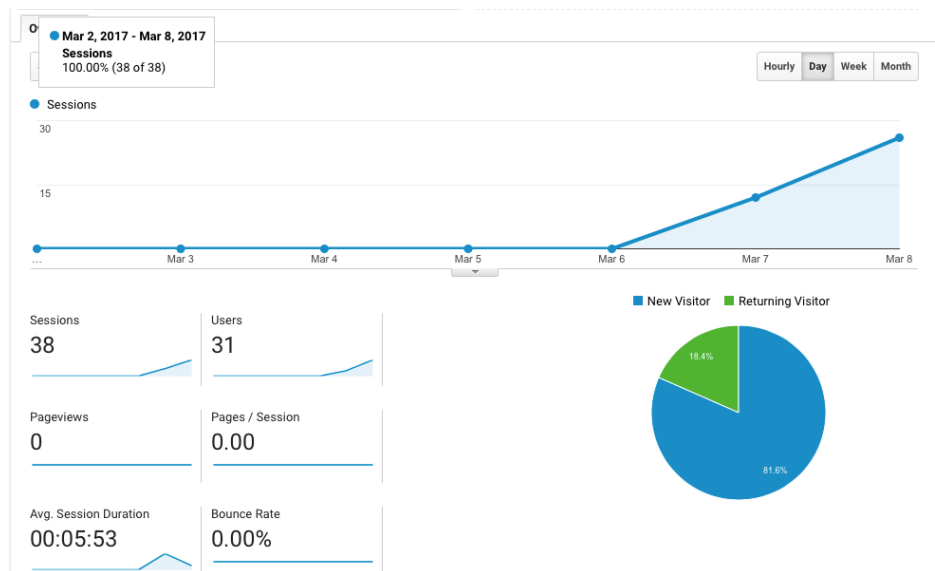

Subjects

Our target group was students, in particular those who attend a university. The average university student leads a very fast paced life and often juggles classes, work, internships and other extra-curricular activities. Due to the abundance of stressors, students frequently have difficulty thoroughly understanding their own goals. Since our app aims to clarify one's thoughts and intentions, university students were an ideal demographic. We asked our friends at UCSD to test our app, as well our high school friends who currently attend different universities. Lastly, we shared our application with the UCSD rowing team. As such, each person who tested the app is guaranteed to be a current student. This way, our target demographic doesn't only consist of students from UCSD but also from a few other universities as well.
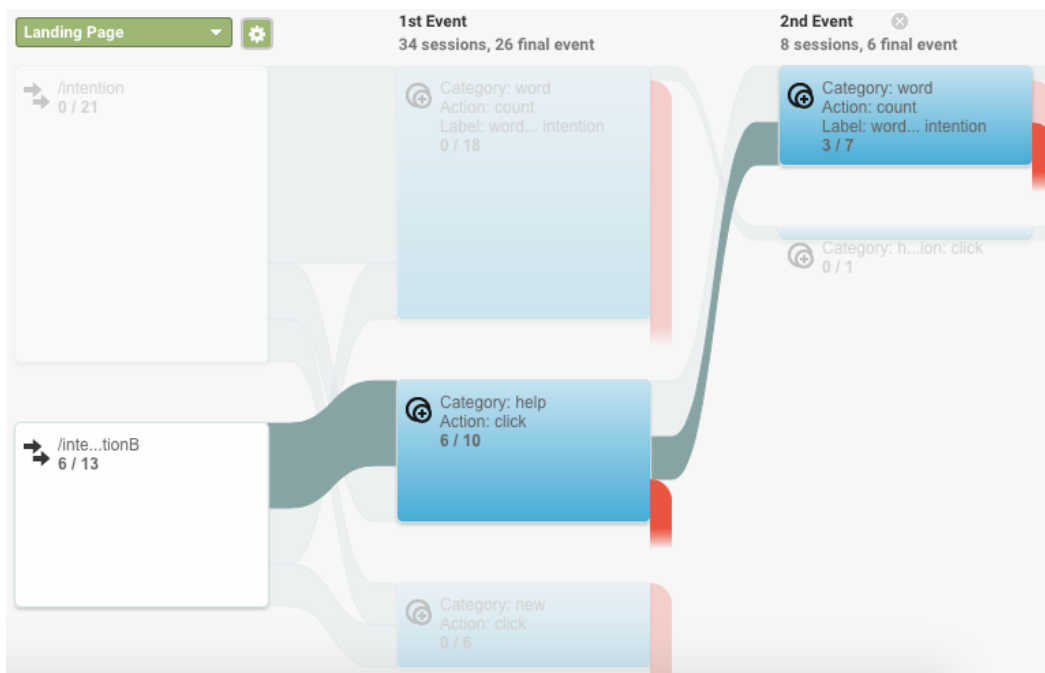
Data

Help Button

Using Google analytics, we recorded clicks on the help button. The two screenshots below from the Event Flow page show the number of users in each condition who clicked 'help', as well as the total number of users who made it to the landing page for each condition.
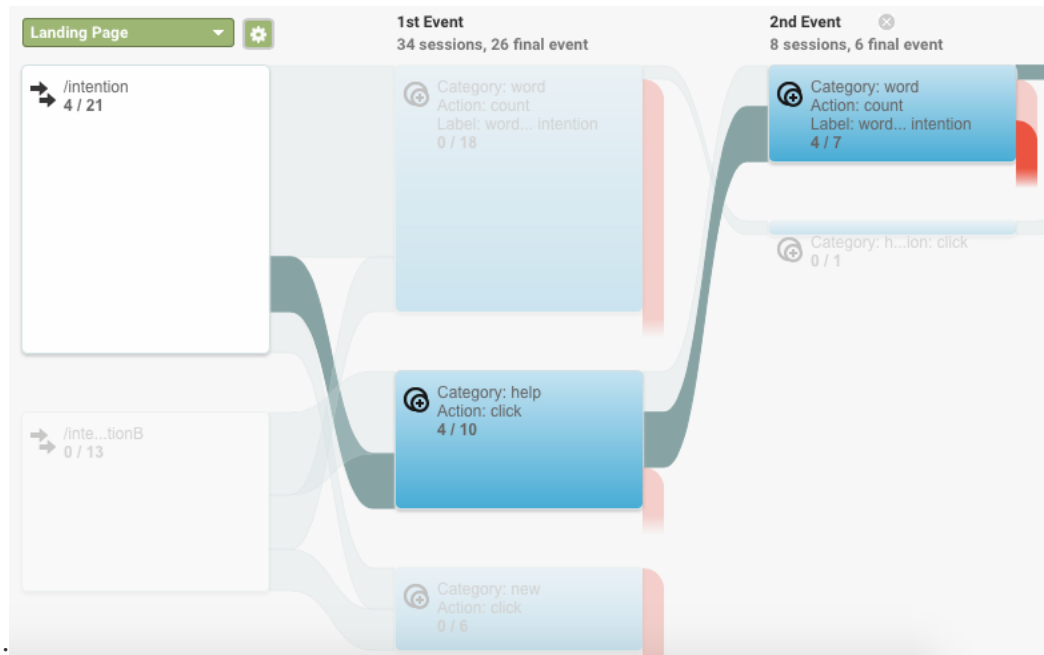
Data for A:

Data for B:

In the following three tables we perform a chi-squared test to determine whether there is a difference in the proportion of users who pressed 'help' across conditions A and B.

Observations	A	B	Total
Pressed Help	**4**	**6**	10
Did not press Help	**17**	**7**	24
Total	21	13	34

Expected	A	B	Total
Pressed Help	6.18	3.82	10
Did not press Help	14.82	9.18	24
Total	21	13	34

(O-E)^2 /E

0.77	1.24
0.32	0.52

2.85

With a critical value of 2.85 and 1 degree of freedom, the p = 0.09. (When run through more precise instruments, the data yield a critical value of 2.8416. This does not change p.) Since p > 0.05, we fail to reject the null hypothesis that merging 'duration' and 'description' does not affect the meaning of 'intention'. While the results not statistically significant, we found the presence of a trend persuasive enough to justify using the format of version B.

Words per field

Due to unforeseen limitations of Google Analytics, we were unable to recover the word counts for specific events. Google Analytics only displays the mean value, with no infrastructure in place to obtain more fine-grained records. As such we cannot recover standard deviation and therefore cannot perform the planned t-test on these data. However, there is still a striking difference between results from the two versions.

Interpreting words per field as a measure of the application's efficacy in promoting user reflection, we hypothesized that version B would be more effective and yield more words per session than version A. This is suggested by our findings shown in the following table, with version B having a mean of 13.6 words/session and version A having a mean of 4.2. This is another factor in our decision to use version B.

	Event Action ?	Landing Screen ? ⊗	Total Events ?	Unique Events ?	Event Value ?	Avg. Value ? ↓
☐			101 % of Total: 48.56% (208)	26 % of Total: 55.32% (47)	415 % of Total: 100.00% (415)	4.11 Avg for View: 2.00 (105.94%)
☐	1. count	on-purpose.herokuapp.com/intentionB	13 (12.87%)	7 (26.92%)	177 (42.65%)	**13.62**
☐	2. count	on-purpose.herokuapp.com/intention	56 (55.45%)	17 (65.38%)	237 (57.11%)	**4.23**

Additional feedback

Many of the users appreciated the animated clouds on our home page.They said that it connected well to the 'mind as sky' theme and added an appreciated dynamism to the interface.

Interpret & Implement

From the results of our analysis, we have found out that the p-value we calculated from the chi-squared test was not significant (p=0.09). Version B is less structured than Version A, meaning that there is more user freedom within B that facilitates for more discussion that is tailored to the users own wants and needs. Version A lays out what our team has deemed significant features of an intention but Version B allows the users to create their own significant features and input whatever they want for the specific intention. This is shown by the significantly increased number of words per intention for B as compared with A.

One interesting thing we noticed is that while version B had more help button presses, both in absolute and proportional terms, it also had a substantially higher average word count. Taken together, these two factors may indicate that clicking help was indeed beneficial for users, i.e. they were confused, they pressed help, and they were able to write more.

Internal & External Validity

Internal Validity : Random assignment of the application ensured that users were only prompted either Version A or B of the 'Create Intentions' Page. The probability was equal that users would either be prompted A or B and thus no problems arose. Also, we made sure that there were no confounds such as maturation effects or history effects since it only takes the user about 5 minutes to test the app.

External : We collected data from over 30 users within our user base. To further improve our external validity, we tested students from other places besides UCSD. We wanted to ensure that our user group was not solely focused on UCSD students, but included students from other universities as well.

Changes Made Based on A/B Testing

We've decided to make a few changes.

1. Instead of having a separate page for Statistics, we are going to include the sentiment analysis of the answers the user typed in the 'Previous Intentions' page. That way, the user can see what their previous intentions are and also see what the mood/emotion of their answers were.
2. We are going to include an 'About' page where we will have a short description of what our app does and why a user would benefit from using our app.

As previously discussed, even though we didn't do a formal test on number of words per question, the number was higher on Version B (13.6) than Version A (4.2) which means that the users on Version B thought more about their intention and delve deeper into their own thought process for that specific intention. This is interesting to us because our app is focussing on getting the users to think critically about their intention and we feel that that was successfully accomplished in both versions — but especially in Version B.

We also noticed that while users did click our 'help' button, through the data, we realized that people were also clicking the button to close the popup window — thus revealing a confounding variable in our results.

What We Learned From Google Analytics

- We learned that we can't see the number of words an individual typed in for a question. We really wanted to run a t-test on the number of words answered per question among the two versions, but we didn't have sufficient data to do so. However, we were able to obtain the average number of words per question for each version.
- We also learned that we can see the trends in user flow and how the users navigate throughout the app. Google Analytics automatically creates a sharp-looking visualization of these trends, which is quite nifty. The common trend was that most users started by creating a new intention (which we had expected).