# Experimentation Summary

Jerry Ye, Jonathan Xu, Vincent Huang

*Jerry, Jonathan, and Vincent were a part of the Spring 2021 edition of Project School, where they explored concepts related to self-experimentation and social accountability. This is a summary of their journey, including the ideas and obstacles they encountered.*

---

## Introduction

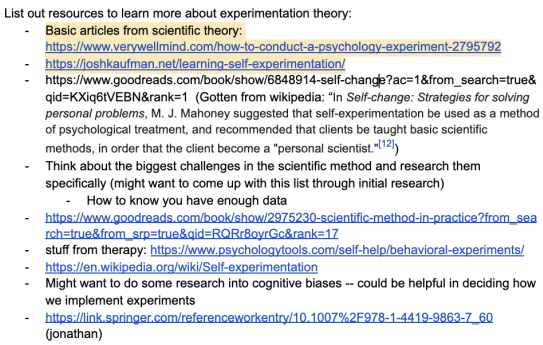
When deciding what project to work on together, all three of us expressed an interest in ideas related to experimentation with our behaviors in the interest of self-improvement. What if you could optimize habits like the way you sleep (a *dependent variable*), by playing around with factors like sleeping position and bedtime and pillow type (*independent variables*) until you found an optimal set of conditions for sleeping? It seemed obvious that doing so would produce massive improvements in quality of life.

We envisioned a community where people experimented frequently and liberally with their lives, recorded and uploaded the results of these experiments, and then shared them with friends. Perhaps if users regularly recorded numerical data about their experiments and independent variables, we could even build computational tools to run inferences and suggest new combinations of variables they hadn't tried. The possibilities seemed endless, so we agreed upon a lofty vision of building the equivalent of a scientist's lab notebook for commonplace activities.

---

## Research

We wanted to start off with doing as much research as possible into both the nature of experimentation itself and optimal ways to design experiments. We were hoping to come up with findings that help us start off with a well designed MVP. Alongside traditional brainstorming, we looked at some books and articles as well:



List out resources to learn more about experimentation theory:

- Basic articles from scientific theory: https://www.verywellmind.com/how-to-conduct-a-psychology-experiment-2795792
- https://joshkaufman.net/learning-self-experimentation/
- https://www.goodreads.com/book/show/6848914-self-change?ac=1&from_search=true&qid=KXiq6tVEBN&rank=1 (Gotten from wikipedia: "In *Self-change: Strategies for solving personal problems*, M. J. Mahoney suggested that self-experimentation be used as a method of psychological treatment, and recommended that clients be taught basic scientific methods, in order that the client become a "personal scientist."[12])
- Think about the biggest challenges in the scientific method and research them specifically (might want to come up with this list through initial research)
    - How to know you have enough data
- https://www.goodreads.com/book/show/2975230-scientific-method-in-practice?from_search=true&from_srp=true&qid=RQRr8oyrGc&rank=17
- stuff from therapy: https://www.psychologytools.com/self-help/behavioral-experiments/
- https://en.wikipedia.org/wiki/Self-experimentation
- Might want to do some research into cognitive biases -- could be helpful in deciding how we implement experiments
- https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-9863-7_60 (jonathan)

From all this information-gathering, we came up a long list of thoughts that we wanted to be prevalent when designing our MVP:

- In traditional experimentation, there are two types of experimentation: exploratory experimentation and confirmatory experimentation. Confirmatory experiments are traditional, they start off with a hypothesis and end with some sort of proof or disproof of the hypothesis. Instead, exploratory experiments are deep dives into some topic without a set goal of what's to be tested for. They avoid bias and tunnel vision, and can lead to results which humans could never come up with a hypothesis for. We wanted to keep the spirit of exploratory experimentation alive, which is one reason why we thought a social aspect to our product would be helpful. People would comment on each others' experiments, providing new ideas and insights.
- By nature, all personal experiments will be scientifically inaccurate because of a low sample size and how some things simply work for some people and don't for others. As an example, someone may experiment and find out 6:00 AM is their ideal wake time while for others it's 9:00 AM. A social aspect to this app will also be helpful in collectively rating experiments that are useful for the most people.
- Being willing to change which independent variables are being tested for is necessary when existing experiment structures fail to provide solid conclusions.
- Lack of motivation to start experimenting and continue experimenting are huge problems
    - Group collaboration and accountability will be helpful in solving this
- Users need to set some sort of timeframe to conclude enough data is collected or else conclusions may never seem like they can be drawn
- Users shouldn't try to optimize for too many independent variables at once or else results may not be able to reveal what a change in certain independent variables led to
- Some experimentations will be done in groups while others will be done individually, but the design will be very similar for both

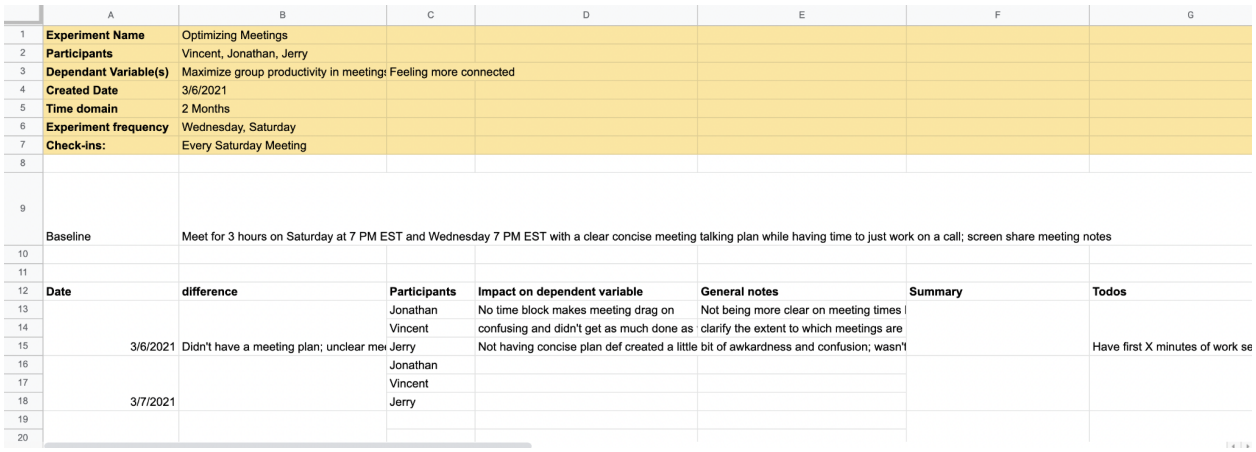
---

## Lots of Iterations

After setting our sights on the goal of building a sort of experimental log for everyday life, we began experimenting with prototypes, starting with very rudimentary spreadsheets. Our initial focus was on giving the user as much flexibility as possible with creating independent and dependent variables. To that end, we created spreadsheets where users could specify any number of independent and dependent variables and record the effects of the independent variables on the dependent ones.

One observation we noticed was that most independent variables have a default value based on whatever “normal life” is like, eg. the default number of meals people eat in a day is three; as a result, whenever users created independent variables they could also keep note of their default values, and then in the future they would only have to log the value of the independent variable if it differed from the norm. We found this idea very promising, because it meant that users could experiment with large numbers of independent variables at once without having to log each variable at every trial.

Here is what the first version of our spreadsheets looked like (the experiment in question is us experimenting with the format of our work meetings to figure out the best way to meet).

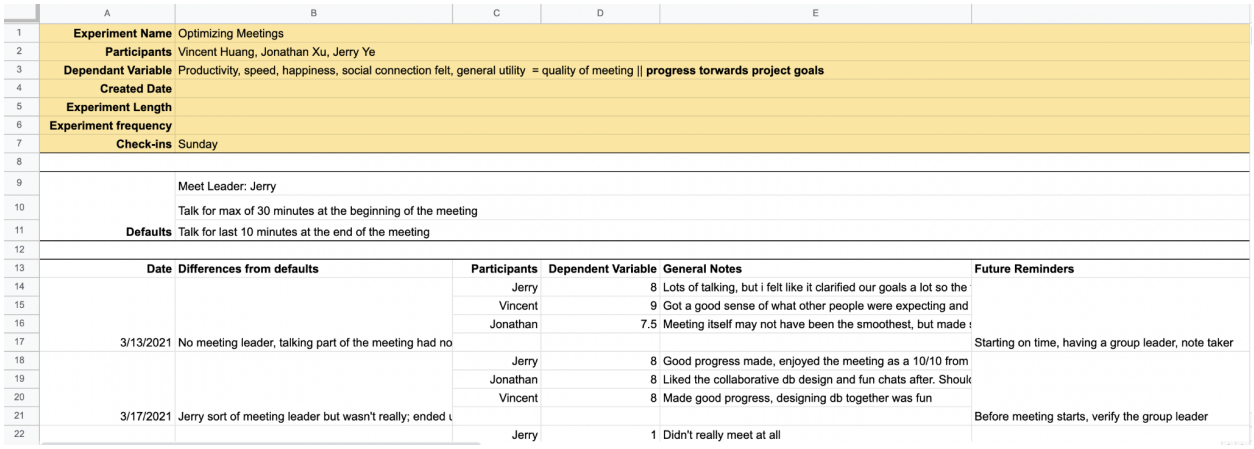


| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | **Experiment Name** | Optimizing Meetings | | | | | |
| 2 | **Participants** | Vincent, Jonathan, Jerry | | | | | |
| 3 | **Dependant Variable(s)** | Maximize group productivity in meetings | Feeling more connected | | | | |
| 4 | **Created Date** | 3/6/2021 | | | | | |
| 5 | **Time domain** | 2 Months | | | | | |
| 6 | **Experiment frequency** | Wednesday, Saturday | | | | | |
| 7 | **Check-ins:** | Every Saturday Meeting | | | | | |
| 8 | | | | | | | |
| 9 | Baseline | Meet for 3 hours on Saturday at 7 PM EST and Wednesday 7 PM EST with a clear concise meeting talking plan while having time to just work on a call; screen share meeting notes | | | | | |
| 10 | | | | | | | |
| 11 | | | | | | | |
| 12 | **Date** | **difference** | **Participants** | **Impact on dependent variable** | **General notes** | **Summary** | **Todos** |
| 13 | | | Jonathan | No time block makes meeting drag on | Not being more clear on meeting times | | |
| 14 | | | Vincent | confusing and didn't get as much done as | clarify the extent to which meetings are | | |
| 15 | 3/6/2021 | Didn't have a meeting plan; unclear mee | Jerry | Not having concise plan def created a little | bit of awkardness and confusion; wasn't | | Have first X minutes of work se |
| 16 | | | Jonathan | | | | |
| 17 | | | Vincent | | | | |
| 18 | 3/7/2021 | | Jerry | | | | |
| 19 | | | | | | | |
| 20 | | | | | | | |

The immediate thing that comes to mind when looking at this prototype is that it is ugly. It also isn’t entirely clear how you’d use the spreadsheet if you wanted to: the most important interaction we wanted to capture was how independent variables affected dependent ones, and the only place this is supported is in some column vaguely named “Impact on dependent variable”, while it’s not clear what other columns like “General notes” and “Summary” are entirely about, or what the distinction between them is.

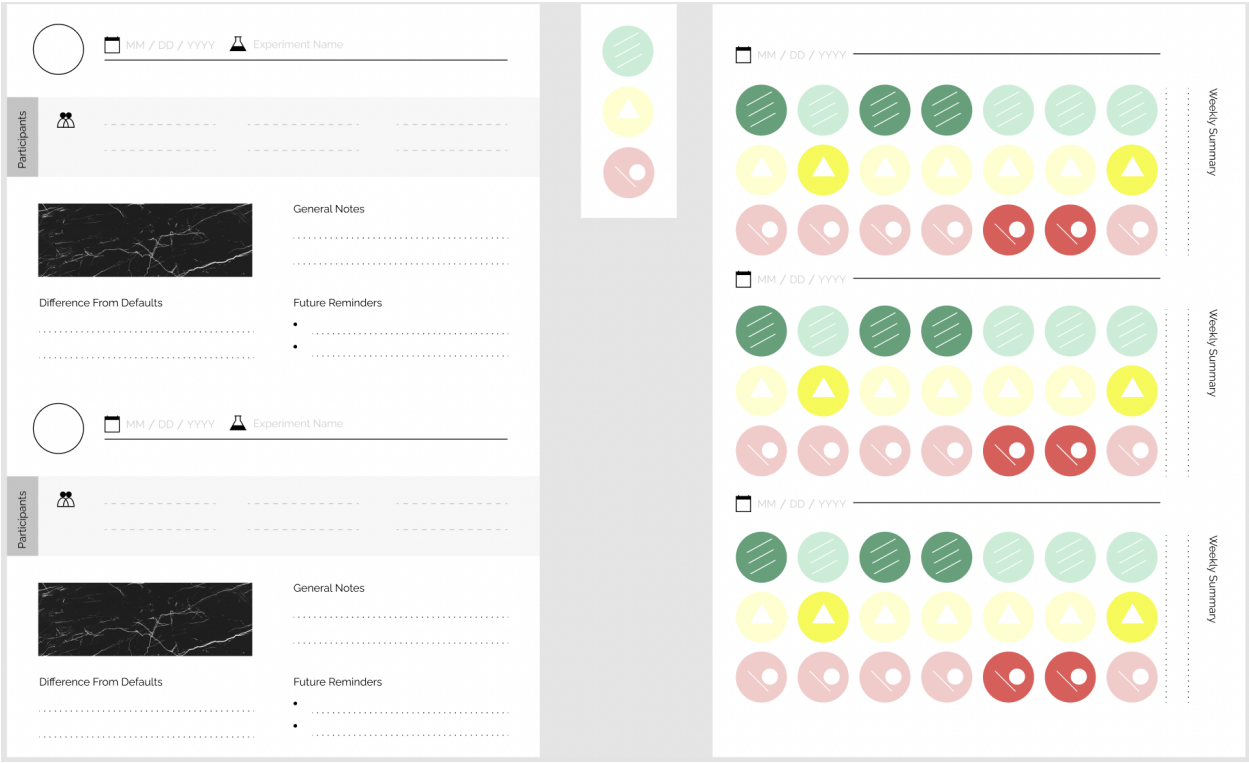
We used this spreadsheet to record information for a few meetings, and quickly realized it was a pain to interact with. It was unclear, however, whether the source of the difficulty was the ugliness mentioned in the previous paragraph, or in the fundamental concept of interacting with a spreadsheet to log a list of variables. As a result, we took a few more iterations to refine our spreadsheets, with help from user testing on our friends and some groups at Project School.

Here is an example of what our spreadsheets ended up looking like by the end of this process. A lot of cells and rows and columns were cleaned up, we settled on measuring a single dependent variable with a single numerical rating, and tried our best to only retain elements that were necessary to our goal.

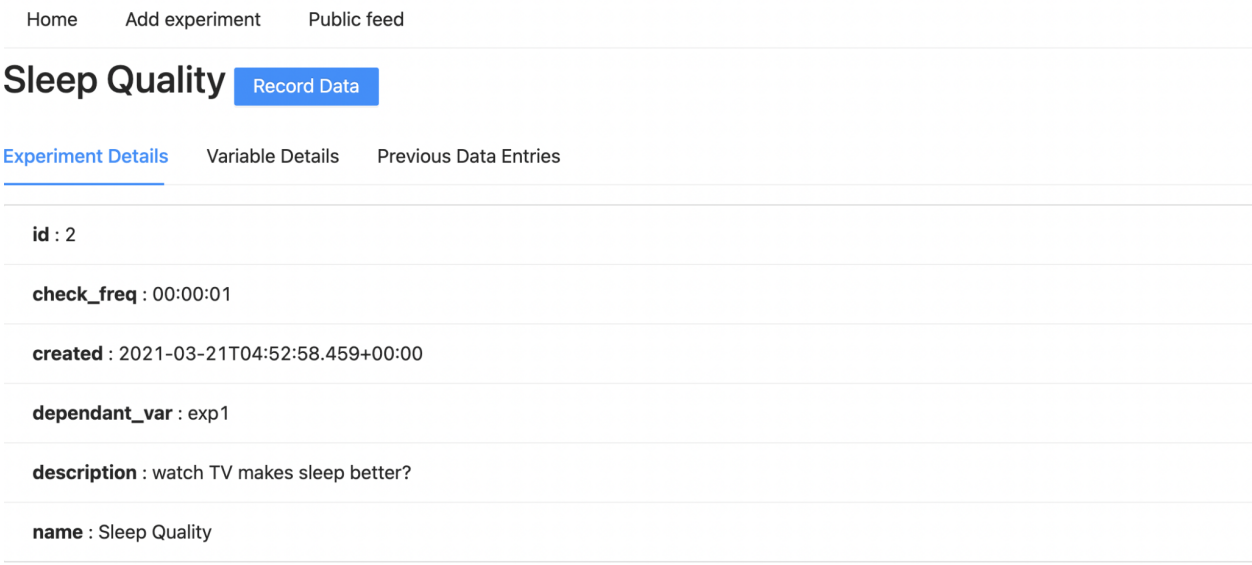


| | | | | | |
|---|---|---|---|---|---|
| **Experiment Name** | Optimizing Meetings | | | | |
| **Participants** | Vincent Huang, Jonathan Xu, Jerry Ye | | | | |
| **Dependant Variable** | Productivity, speed, happiness, social connection felt, general utility = quality of meeting \|\| **progress torwards project goals** | | | | |
| **Created Date** | | | | | |
| **Experiment Length** | | | | | |
| **Experiment frequency** | | | | | |
| **Check-ins** | Sunday | | | | |
| | | | | | |
| | Meet Leader: Jerry | | | | |
| | Talk for max of 30 minutes at the beginning of the meeting | | | | |
| **Defaults** | Talk for last 10 minutes at the end of the meeting | | | | |
| | | | | | |
| **Date** | **Differences from defaults** | **Participants** | **Dependent Variable** | **General Notes** | **Future Reminders** |
| | | Jerry | 8 | Lots of talking, but i felt like it clarified our goals a lot so the | |
| | | Vincent | 9 | Got a good sense of what other people were expecting and | |
| | | Jonathan | 7.5 | Meeting itself may not have been the smoothest, but made | |
| 3/13/2021 | No meeting leader, talking part of the meeting had no | | | | Starting on time, having a group leader, note taker |
| | | Jerry | 8 | Good progress made, enjoyed the meeting as a 10/10 from | |
| | | Jonathan | 8 | Liked the collaborative db design and fun chats after. Should | |
| | | Vincent | 8 | Made good progress, designing db together was fun | |
| 3/17/2021 | Jerry sort of meeting leader but wasn't really; ended | | | | Before meeting starts, verify the group leader |
| | | Jerry | 1 | Didn't really meet at all | |

This was somewhat easier to use, and as you can see we actually recorded a few more weeks' worth of meeting logs in this variant of the spreadsheet, and found it somewhat more enjoyable. But it was still a far cry from the ease and convenience of writing things down on paper, and people whom we asked to try out our sheets generally had a difficult time finding motivation to record information regularly. So we began brainstorming alternative mediums to spreadsheets. One promising idea was a printable book with colored stickers to indicate level of progress:




Due to an upcoming demo deadline and our troubles with finding a notebook maker that let us design custom pages and stickers, we started building a webapp version of our spreadsheets:



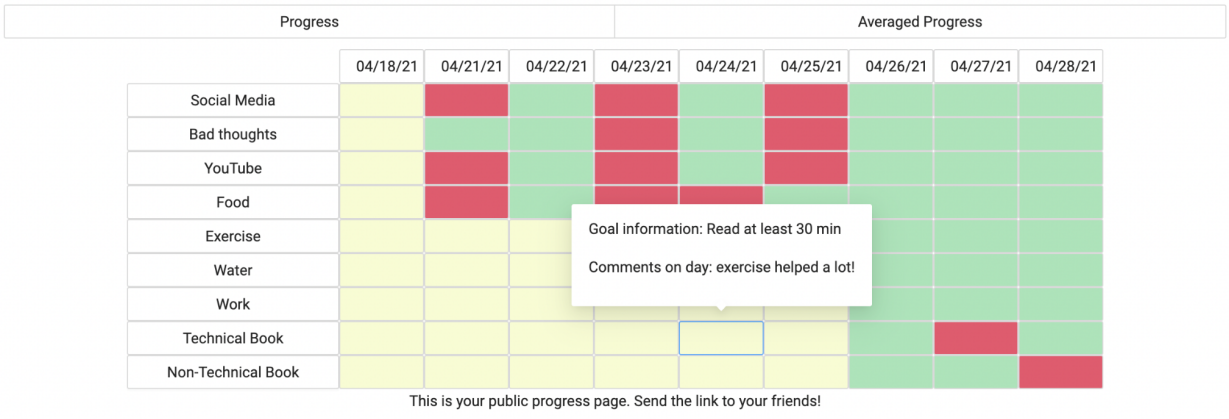

However, by this point our ongoing experiment of exploring different meeting formats had begun to fizzle out, and it became apparent even before we finished implementing core website features that this would not be a worthwhile use of time. We noticed that the variables we were mixing up, like the meeting leader and time allocation and such, didn't seem to have any discernible impact on how well the meetings were going, or at the very least any impact was so minimal it'd likely take hundreds of trials to notice. Our original vision was to record large

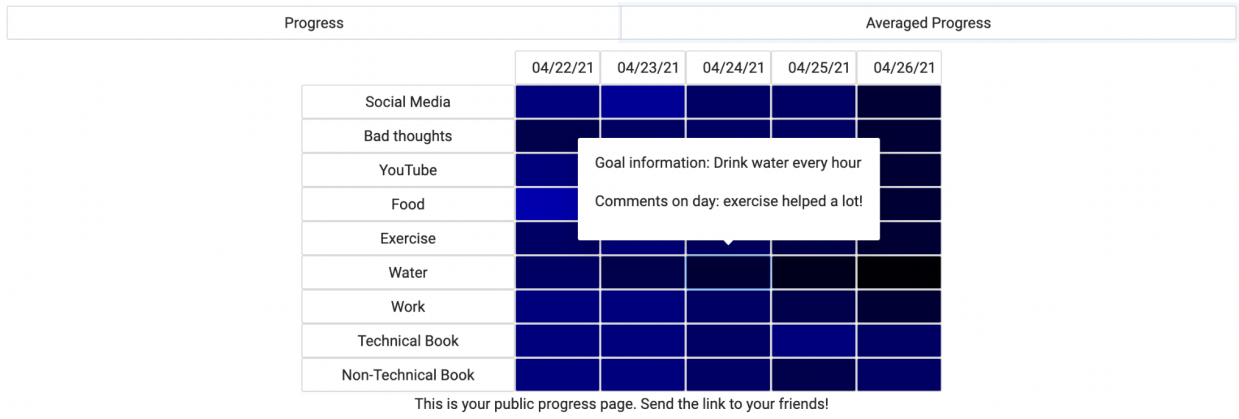
numbers of variables and eventually learn the optimal combinations of them, but if we couldn't find immediate correlations in the obviously important variables then it made even less sense to record the less important variables. Indeed, the primary determinants of how well our meetings went were factors like how happy or tired we happened to be on any particular day, which didn't seem particularly meaningful to log, and were far too strongly correlated with factors outside of our control anyway. In principle we could've begun recording these variables too and built a model which would learn to discern the effects of variables within our control and ignore the effects of variables that weren't, but ideas along these lines seemed increasingly infeasible.

Essentially, the product we were building was not practical because humans don't think in terms of optimizing the impact of every last variable in each of their goals; only machines can act with this level of rigor and precision. Even the most accessible version of our product, the printable book, would likely suffer from the same issues. As a result, it became clear that a usable version of our product would not revolve around logging measuring and logging the values of various variables; we would have to simplify our vision significantly.

One thing we continued thinking a lot about was the unreasonable effectiveness of paper. In many ways, paper is still the gold standard for note-taking, because it is easy to view and write in, and anything written down is immediately persistent. The shortcomings of paper are shareability and fancy formatting like variety in font style or color, so ideally a digital information tracker would preserve the low activation presented by paper while providing some kind of visualization in a way that paper cannot.

This led to our final idea, a website where users specify a number of goals, and every day they mark whether a goal was successful or not, and any other comments about the day. The site then displays simple and colorful views of their goal progress over time, and when users notice any particularly successful or unsuccessful days they can review the comments from those days. In essence we shifted the focus of the product away from interaction between variables and towards visual display of information. This is much simpler of a product than what we originally envisioned, but at least it's something that feels more usable.






We are now in the testing phase of this evolution of our product, and will circle back to write about the usefulness and efficacy of our new design. With the pivot from obsession over detail and quantifiable metrics to low-activation logging and trend discovery, we hope we can discover much more about the science of optimizing our lives.