Evaluación de Interfaces de Usuario

Clase 3
ECI 2024

Carla F. Griggio, tenure-track assistant professor cfg@cs.aau.dk
Department of Computer Science, Human-Centered Computing (HCC)
Aalborg University, Copenhagen

AALBORG UNIVERSITY

ECI37

.UBA

Comunicación durante el curso

Para dudas, compartir contenido interesante, y catársis sobre interfaces mal diseñadas

Grupo de Slack de SIGCHI LAIHC
Latin American HCI Community.
Únanse al channel **#eci-uba-2024**
para mantenernos comunicados.

Clase 3

- Intro a evaluaciones empíricas
- Diseño de experimentos controlados
- (si llegamos) Estudios de campo

Métodos de evaluación empíricos (User studies)

(Ojo con la traducción)

Por lo que estuve viendo, la traducción más común de “user study” es “estudio **de** usuarios”.

Técnicamente, sí, estamos estudiando “**a**" los usuarios al observar su comportamiento mientras interactúan con nuestro sistema.

Pero en términos de evaluación, estamos evaluando el sistema, no a los usuarios!

En esta clase voy a preferir decir “**estudio con usuarios**” para evitar confusiones. Es importante al momento de empezar un estudio con usuarios aclararles que el objetivo del estudio es evaluar el sistema, y no a su rendimiento.

Para qué involucrar usuarios en una evaluación?
Cuándo nos quedan cortas las evaluaciones analíticas?

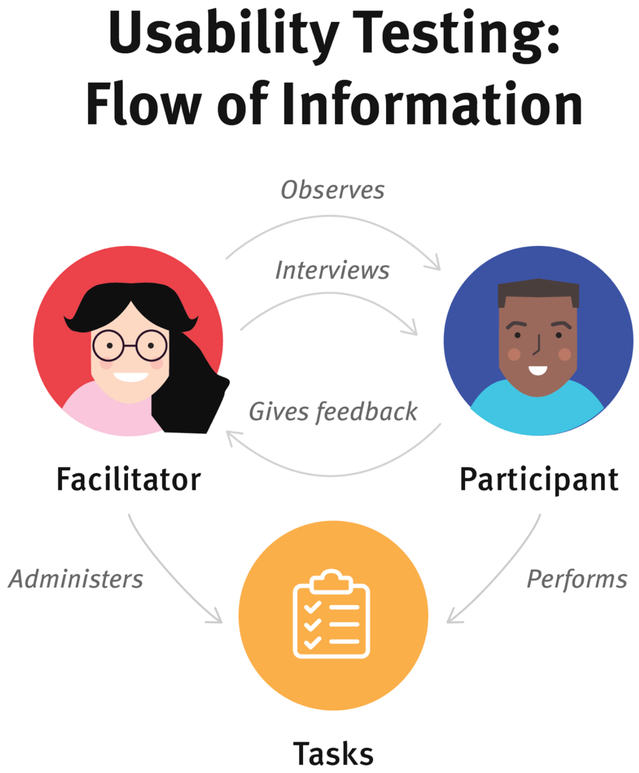
Los estudios con usuarios pueden ayudar en varios tipos de objetivos de evaluación y/o investigación

1) Evaluar la usabilidad de un sistema

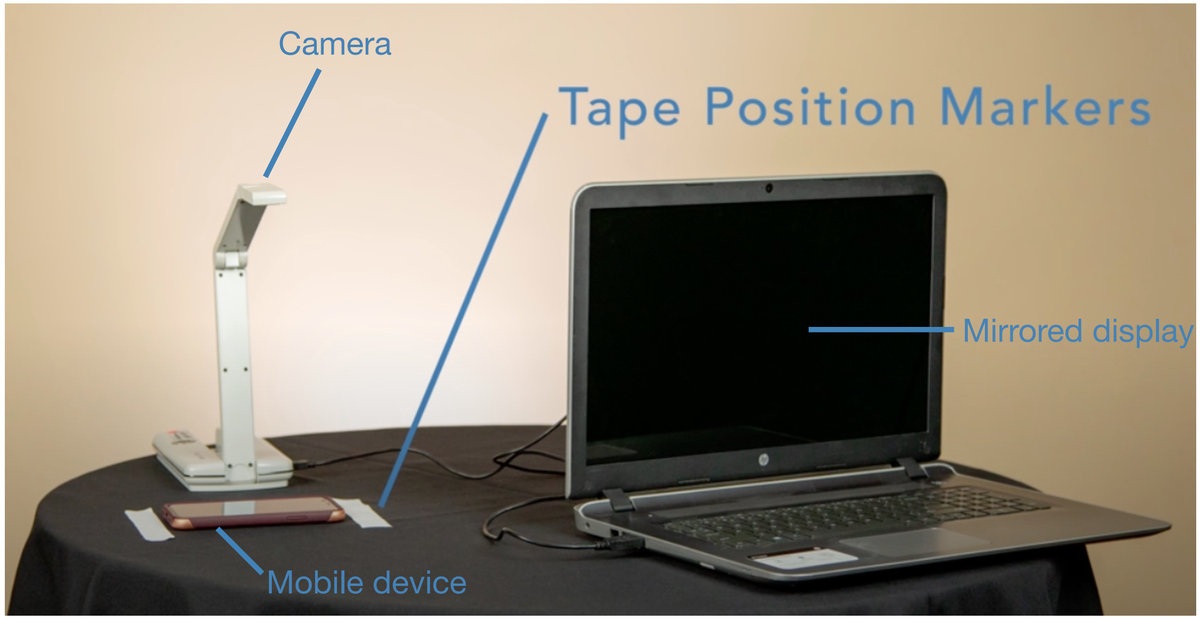
La usabilidad es una cualidad de un sistema que se refiere a qué tan **eficiente**, **eficaz** y **satisfactorio** es usarlo.

El objetivo está muy enfocado **en encontrar problemas de usabilidad** según esa definición.

No hay un criterio standard para concluir si un sistema “es” o “no es” usable. Queda a criterio del diseñador/desarrollador/compañía, y sobre todo, del objetivo y contexto de uso de un sistema.



1) Evaluar la usabilidad de un sistema



https://www.youtube.com/watch?v=NdqTbpyvalg

Estudio de usabilidad
Con/Sin “Think Alound Protocol”
(Pensar en voz alta)

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

Physical Demand — How physically demanding was the task?
Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High

System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Cuestionarios estandarizados
(SUS, NASA-TLX)

1) Entender las necesidades de los usuarios

Emoji Accessibility for Visually Impaired People

Garreth W. Tigwell
Rochester Institute of Technology
Rochester, NY, USA
garreth.w.tigwell@rit.edu

Benjamin M. Gorman
Bournemouth University
Bournemouth, England, UK
bgorman@bournemouth.ac.uk

Rachel Menzies
University of Dundee
Dundee, Scotland, UK
r.menzies@dundee.ac.uk

ABSTRACT
Emoji are graphical symbols that appear in many aspects of our lives. Worldwide, around 36 million people are blind and 217 million have a moderate to severe visual impairment. This portion of the population may use and encounter emoji, yet it is unclear what accessibility challenges emoji introduce. We first conducted an online survey with 58 visually impaired participants to understand how they use and encounter emoji online, and the challenges they experience. We then conducted 11 interviews with screen reader users to understand more about the challenges reported in our survey findings. Our interview findings demonstrate that technology is both an enabler and a barrier, emoji descriptors can hinder communication, and therefore the use of emoji impacts social interaction. Using our findings from both studies, we propose best practice when using emoji and recommendations to improve the future accessibility of emoji for visually impaired people.

and are used by politicians and government bodies [36, 55], travel companies [54], media outlets, and public figures (e.g., singer Katy Perry who has one of the largest Twitter followings [51]). Emoji have even been discussed within official court transcripts [35], and resulted in convictions [23].

People interpret emoji differently, and emoji design variations across different platforms (e.g., iOS vs Android) can exacerbate misunderstandings [45, 64]. Furthermore, emoji are often used beyond their original intended meaning, which adds another layer of complexity to disambiguating the intended use of an emoji [64, 74]. Prior research on emoji has largely focused on those with typical vision. However, it is estimated that 36 million people worldwide are blind and 217 million have a moderate to severe visual impairment [73]. Prior work highlighted challenges visually impaired people face when using technology [7] and social media [22, 49]. However, it is not clear what accessibility challenges occur with emoji.

Poor Use in Context: Our participants highlighted that emoji used in different contexts can lead to specific challenges. Decorative emoji, e.g. emoji in usernames on social media, caused challenges as many decorative emoji could be announced by a screen reader. An example of this is shown in Figure 2.A.

P7: *"Try listening to 'cat with heart shaped eyes fireworks sparkles watermelon kissing face flag of Andorra' a few times in a row and you get the frustration."*

Visual Design: Emoji design also caused challenges for 22% of the participants. For participants who had some residual vision, this was often related to the use of colour such as P6 who described that *"the colors of the heart [emoji] can be too similar."*. For blind participants, differences between design of the visual emoji and the description were challenging:

P28: *"Some emoji [are] useless or just have a bad design (I was told the 'pray' emoji [🙏] is actually a 'high five')."*

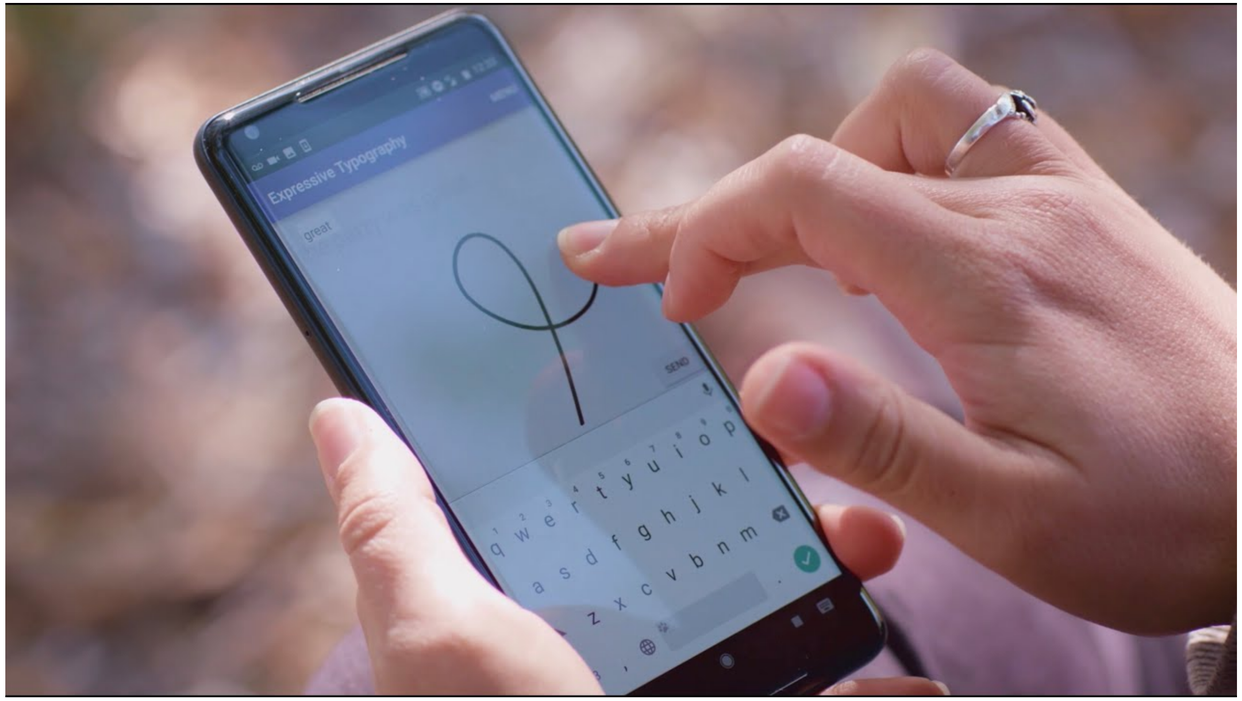
Misunderstanding: This relates to the use of visual representations of things that blind users had not experienced. This sometimes made it difficult to select an emoji.

P38: *"...I entered the word 'happy', and it suggested many faces, which were all described to me; however, as I have never had vision, I was unable to know which face was the most appropriate for my situation."*

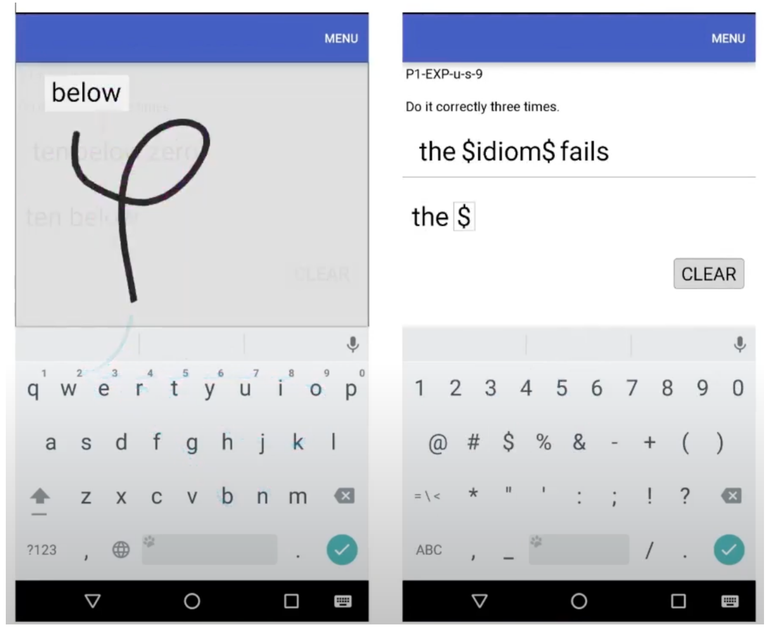
Entrevistas semi-estructuradas

Cuestionarios

3) Comparar alternativas de diseño



Jessalyn Alvina, Carla F. Griggio, Xiaojun Bi, and Wendy E. Mackay. 2017. CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboard (UIST '17). https://doi.org/10.1145/3126594.3126639

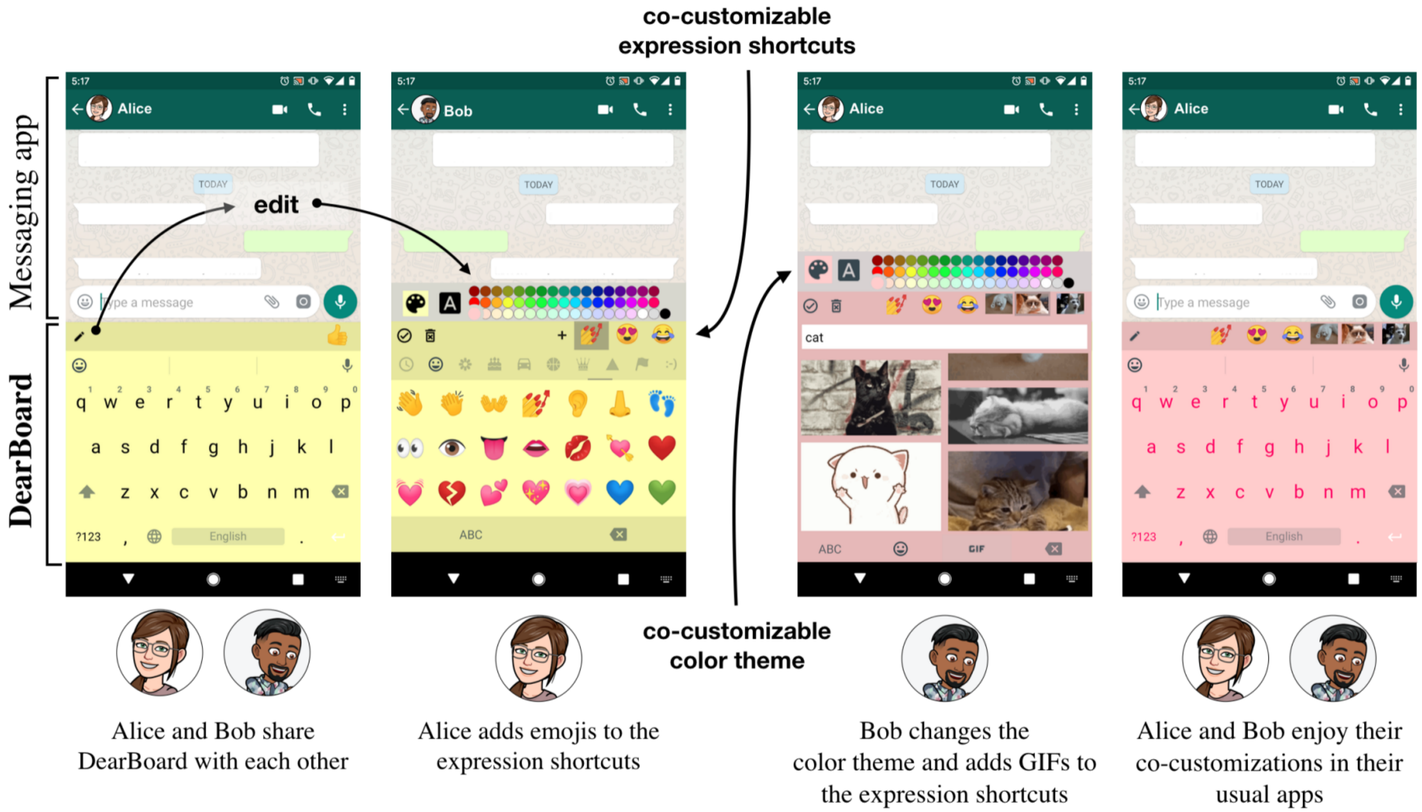


3.3 seconds 6.3 seconds

Tiempo que lleva tipear una palabra en negrita

Experimentos

4) Entender cómo un sistema se usa en el mundo real



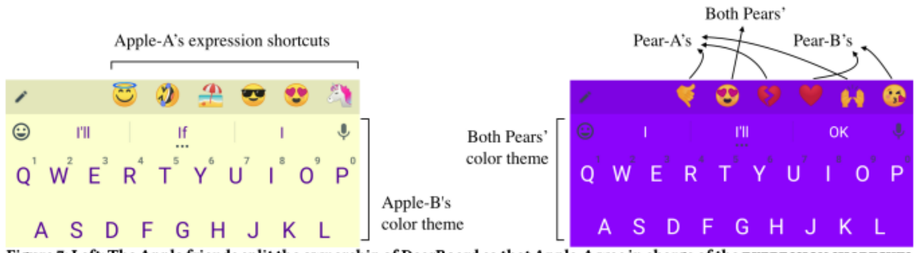


Figure 7: Left: The Apple friends split the ownership of DearBoard so that Apple-A was in charge of the EXPRESSION SHORTCUTS and Apple-B in charge of the COLOR THEME. Right: The Pear couple split the ownership of the EXPRESSION SHORTCUTS to have some for Pear-A and some for Pear-B, and some for both; their COLOR THEME was purple, their favorite color.

Figure 5: The Peach friends typed on an "invisible" keyboard for eight days straight after one of them chose the same color for the background and text as a joke.

Estudios de campo

Carla F. Griggio, Arissa J. Sato, Wendy E. Mackay, and Koji Yatani. 2021. Mediating Intimacy with DearBoard: a Co-Customizable Keyboard for Everyday Messaging (CHI '21). https://doi.org/10.1145/3411764.3445757

5) Descubrir nuevas oportunidades de diseño / innovación

Emoji Accessibility for Visually Impaired People

Garreth W. Tigwell
Rochester Institute of Technology
Rochester, NY, USA
garreth.w.tigwell@rit.edu

Benjamin M. Gorman
Bournemouth University
Bournemouth, England, UK
bgorman@bournemouth.ac.uk

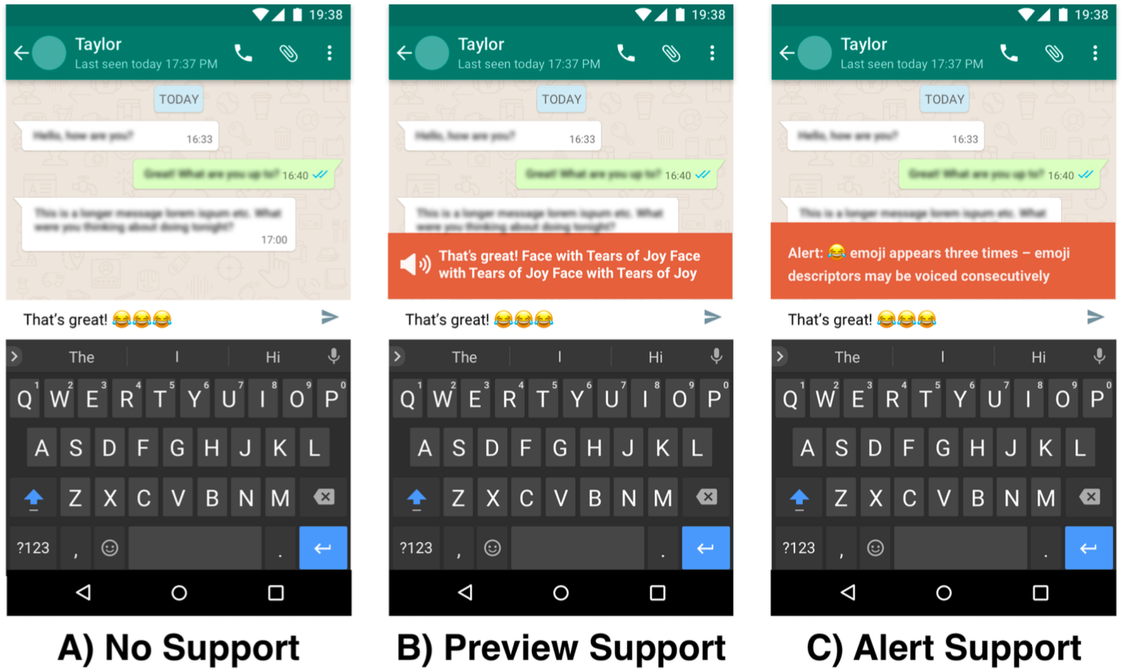
Rachel Menzies
University of Dundee
Dundee, Scotland, UK
r.menzies@dundee.ac.uk

ABSTRACT

Emoji are graphical symbols that appear in many aspects of our lives. Worldwide, around 36 million people are blind and 217 million have a moderate to severe visual impairment. This portion of the population may use and encounter emoji, yet it is unclear what accessibility challenges emoji introduce. We first conducted an online survey with 58 visually impaired participants to understand how they use and encounter emoji online, and the challenges they experience. We then conducted 11 interviews with screen reader users to understand more about the challenges reported in our survey findings. Our interview findings demonstrate that technology is both an enabler and a barrier, emoji descriptors can hinder communication, and therefore the use of emoji impacts social interaction. Using our findings from both studies, we propose best practice when using emoji and recommendations to improve the future accessibility of emoji for visually impaired people.

and are used by politicians and government bodies [36, 55], travel companies [54], media outlets, and public figures (e.g., singer Katy Perry who has one of the largest Twitter followings [51]). Emoji have even been discussed within official court transcripts [35], and resulted in convictions [23].

People interpret emoji differently, and emoji design variations across different platforms (e.g., iOS vs Android) can exacerbate misunderstandings [45, 64]. Furthermore, emoji are often used beyond their original intended meaning, which adds another layer of complexity to disambiguating the intended use of an emoji [64, 74]. Prior research on emoji has largely focused on those with typical vision. However, it is estimated that 36 million people worldwide are blind and 217 million have a moderate to severe visual impairment [73]. Prior work highlighted challenges visually impaired people face when using technology [7] and social media [22, 49]. However, it is not clear what accessibility challenges occur with emoji.



Carla F. Griggio, Benjamin M. Gorman, and Garreth W. Tigwell. 2024. Party Face Congratulations! Exploring Design Ideas to Help Sighted Users with Emoji Accessibility when Messaging with Screen Reader Users. Proc. ACM Hum.-Comput. Interact. 8, CSCW1, Article 175 (April 2024), 31 pages. https://doi.org/10.1145/3641014

Experimentos controlados

(en realidad, decir experimento controlado es redundante, porque todos los experimentos son controlados, así que a partir de ahora, se dice “experimento" y nada más :D)

Experimentos

Un experimento, en HCI, es un tipo de estudio con usuarios que **compara** dos o más alternativas de diseño respecto a las mismas métricas.

Hay dos tipos de métricas que suelen medirse para comparar diseños:

Métricas de rendimiento (performance metrics): datos fácticos y que capturan aspectos del comportamiento o acciones del usuario. Estos datos son "**observados**" por el investigador. Por ejemplo, el número de clicks que se necesitan para terminar una tarea, o el tiempo que toma terminar una tarea.

Métricas de experiencia de usuario (experience metrics): datos subjetivos sobre la satisfacción y experiencia del usuario. Estos datos son **reportados** por los usuarios. Por ejemplo: qué tan fácil o frustrante les resultó cada diseño en una escala del 1 al 5.

En ambos casos, son métricas cuantitativas. No confundir cuantitativo con "objetivo"!

Métricas: datos cuantitativos

Métricas continuas: pueden tomar cualquier valor numérico. Ejemplos:

- el tiempo que le toma al usuario terminar una tarea.
- cuántos píxeles scrollea el usuario antes de abandonar una página.
- el nivel de zoom en una interfaz zoomeable (por ejemplo, Google Maps).

Métricas discretas: pueden tomar valores numéricos específicos, típicamente enteros. Ejemplos:

- Cantidad de errores mientras el usuario completaba una tarea
- Cantidad de clicks hasta terminar una tarea

Métricas ordinales: los valores posibles siguen un orden específico. Ejemplos:

- Preferencia entre los distintos diseños evaluados
- Qué tan satisfactorio/frustrante/interesante/etc. le pareció al usuario cada diseño en una escala de Likert (e.g., seleccionar un nivel de satisfacción entre 1 y 5, donde 1=Nada Satisfactorio, 5= Muy Satisfactorio)

Una vez que el experimento termina de ejecutarse con todos los participantes, el **análisis de datos es estadístico**, con estadísticas descriptivas (e.g., el tiempo promedio que tomó terminar una tarea con cada diseño) o tests estadísticos (e.g., un t-test/prueba de t de Student para testear una hipótesis sobre qué diseño es más eficiente para una tarea en particular)

Datos qualitativos

En un experimento también podemos recolectar datos qualitativos, que no pueden ser medidos numéricamente.

Datos nominales o categóricos:

- El diseño favorito del participante (A o B)
- Ocupación del participante (Trabajo full time / part time / estudiante / desempleado / otro)

Preguntas abiertas: pueden hacerse con cuestionarios o entrevistas durante / después del experimento:

- Por qué preferís el diseño A?
- Qué te pareció lo más difícil de entender del diseño B?

Notas de observaciones tomadas por los investigadores:

- El participante 21 se queja en voz alta cada vez que le toca usar el diseño B

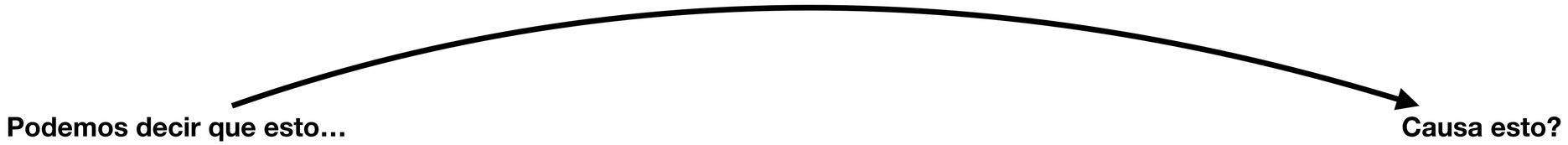
Para analizar datos nominales/categóricos, también podemos usar estadísticas descriptivas. Ejemplo: cuántas personas desempleadas participaron del estudio, qué porcentaje eligió el diseño A como su favorito.

Para notas y respuestas a preguntas abiertas, se usan técnicas de análisis interpretativo para buscar patrones y crear categorizaciones de manera estructurada. Ejemplos: open/axial coding, diagramas de afinidad (affinity diagram), análisis temático (thematic analysis).

Experimentos

Un experimento, en HCI, es un tipo de estudio con usuarios que **compara** dos o más alternativas de diseño respecto a las mismas métricas.

Los experimentos buscan establecer una relación causa-efecto entre un diseño en particular y una métrica, en el contexto de una tarea en particular.



Un diseño

Botones más grandes / más chicos
Input de texto / input de voz
Menú en el borde superior / menú en el borde derecho
etc

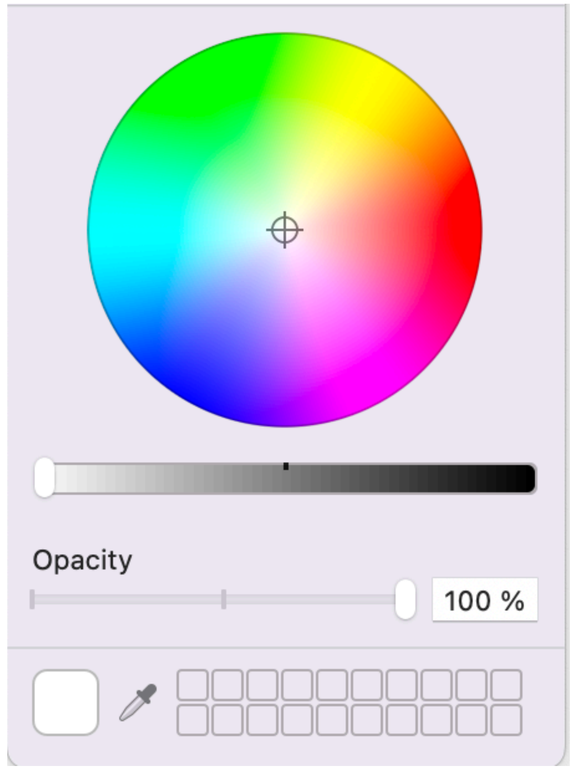
Una métrica

Tiempo para terminar una tarea
Número de errores
Nivel de satisfacción/frustración
etc

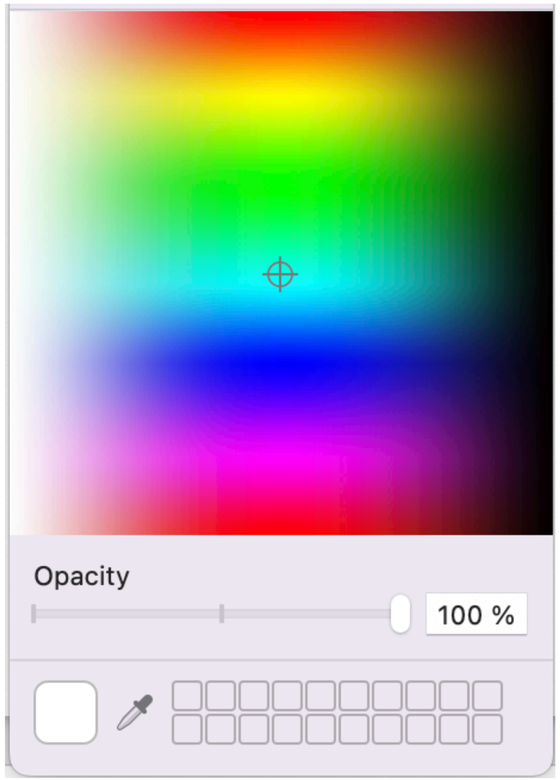
Diseño experimental

Experimental design

Queremos saber si el Color Picker A es mejor para seleccionar colores que el Color Picker B



Color Picker A



Color Picker B

Diseño experimental

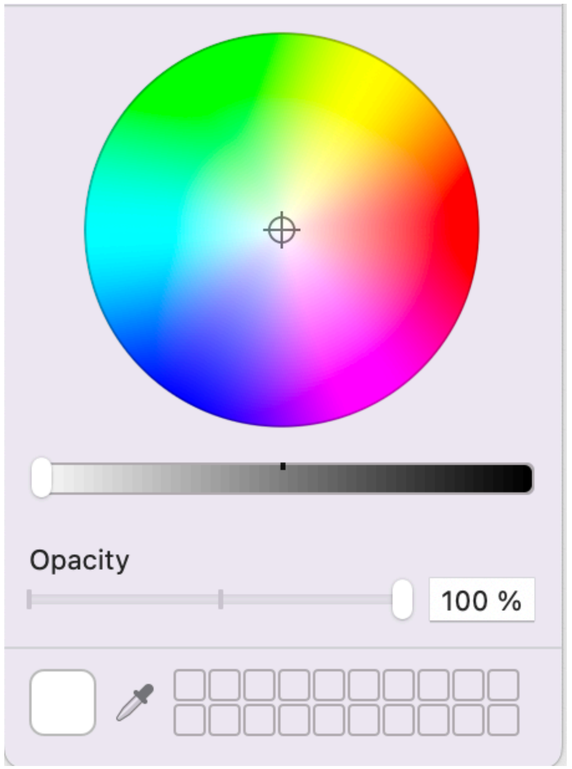
Experimental design

Queremos saber si el Color Picker A es mejor para seleccionar colores que el Color Picker B.

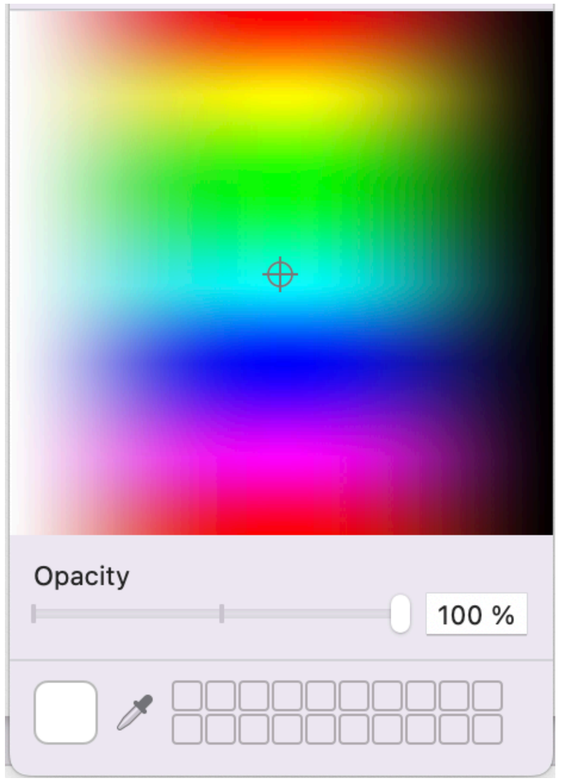
Si “mejor" = “más preciso”:

Cuál es la relación entre el color picker y la precisión con la que un usuario puede seleccionar un color.

Hipótesis 1: El Color Picker A permite seleccionar colores con más precisión que el Color Picker B



Color Picker A



Color Picker B

Diseño experimental

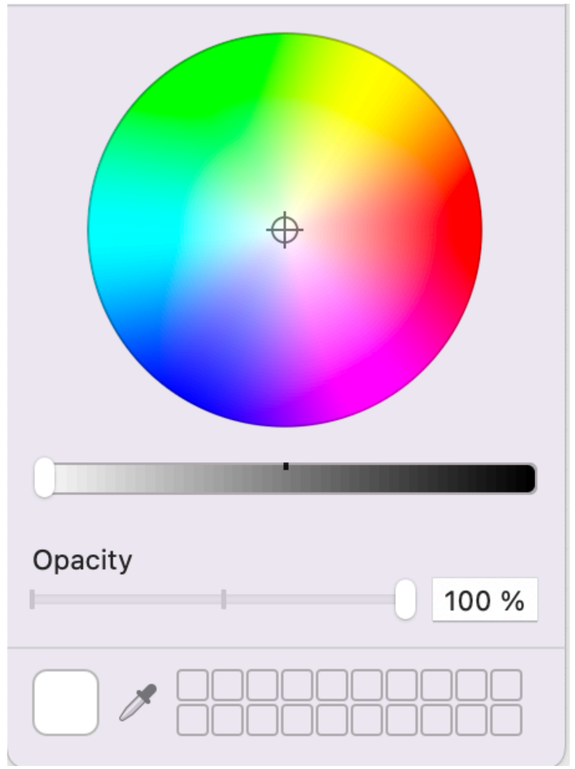
Experimental design

Queremos saber si el Color Picker A es mejor para seleccionar colores que el Color Picker B.

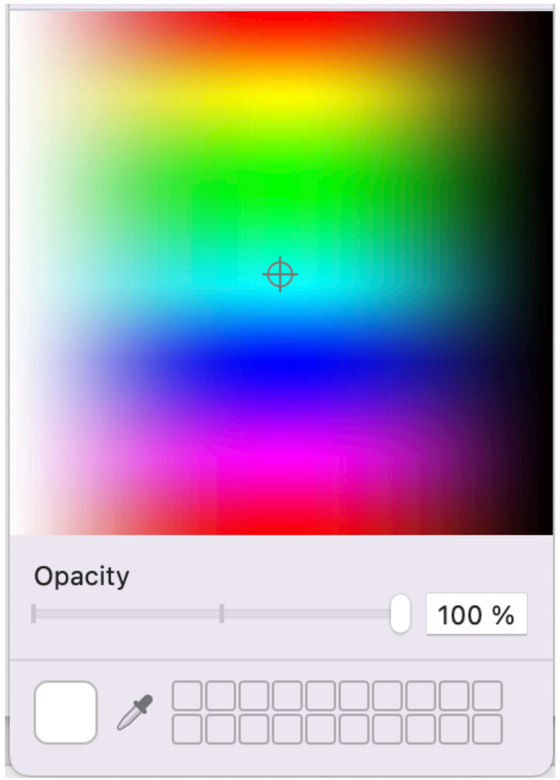
Hipótesis 1: El Color Picker A permite seleccionar colores con más precisión que el Color Picker B

Hipótesis 2: El Color Picker A permite seleccionar colores más rápido que el Color Picker B

La tarea? Qué es lo que tiene que hacer el usuario de manera que podamos registrar las métricas de precisión y velocidad de selección?



Color Picker A



Color Picker B

Diseño experimental

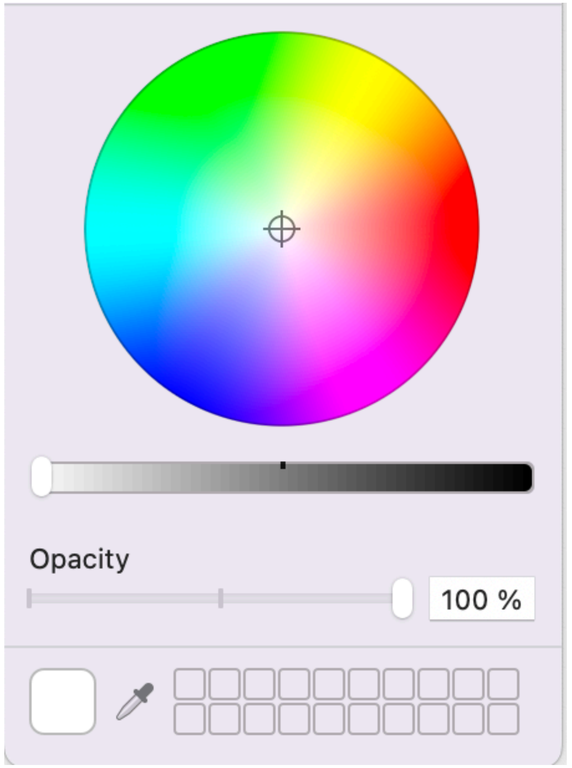
Experimental design

Queremos saber si el Color Picker A es mejor para seleccionar colores que el Color Picker B.

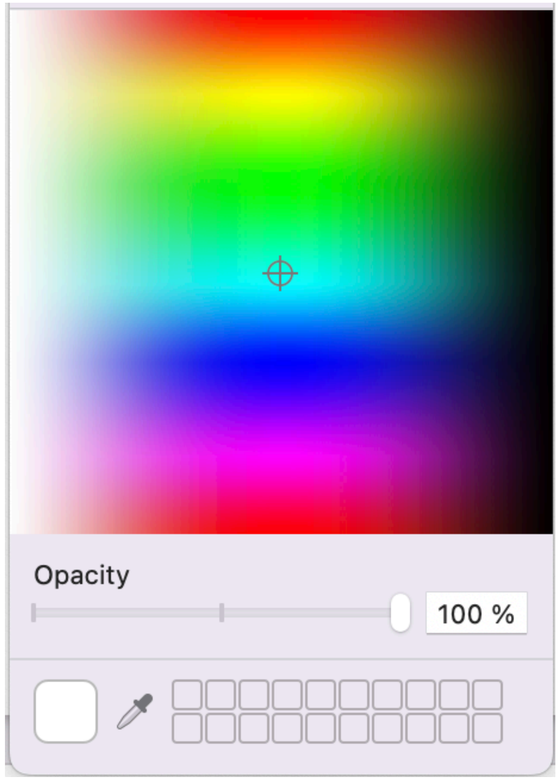
Hipótesis 1: El Color Picker A permite seleccionar colores con más precisión que el Color Picker B

Hipótesis 2: El Color Picker A permite seleccionar colores más rápido que el Color Picker B

Tarea: "Arrastrá tu cursor para encontrar el color objetivo. Intentá hacerlo lo más rápido y preciso posible.”



Color Picker A



Color Picker B

Diseño experimental

Experimental design

Trial: Cada una de las veces que un participante ejecuta la tarea.

Variable dependiente (DV): una métrica. Son variables que “dependen” del usuario.

Variable independiente (IV): aspectos del diseño o estudio que controla el investigador. Por ejemplo: la variable “Color Picker” varía entre A y B cada vez que el usuario repite la tarea.

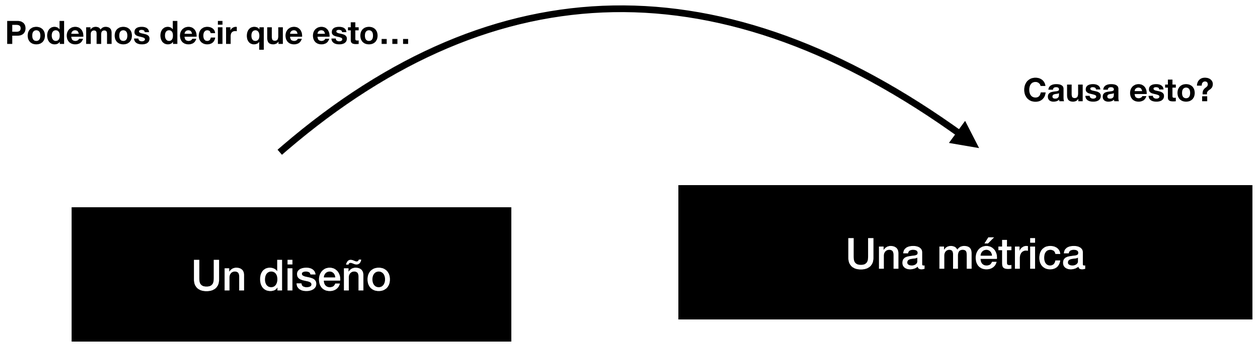
Diseño experimental

Experimental design

Trial: Cada una de las veces que un participante ejecuta la tarea.

Variable dependiente (DV): una métrica. Son variables que “dependen” del usuario.

Variable independiente (IV): aspectos del diseño o estudio que controla el investigador. Por ejemplo: la variable “Color Picker” varía entre A y B cada vez que el usuario repite la tarea.



Botones más grandes / más chicos
Input de texto / input de voz
Menú en el borde superior / menú en el borde derecho
etc

Variable independiente

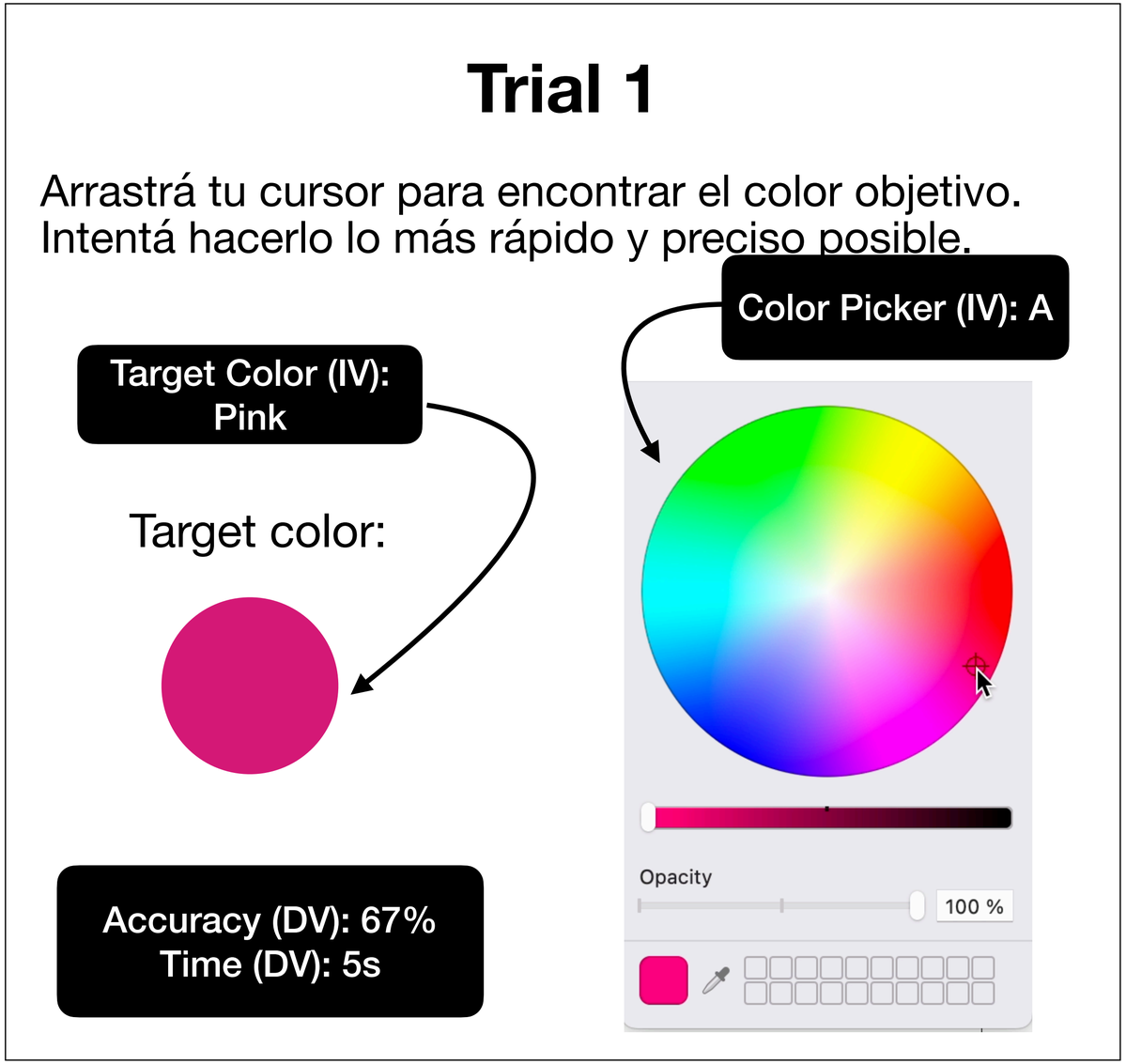
Tiempo para terminar una tarea
Número de errores
Nivel de satisfacción/ frustración
etc

Variable dependiente

Diseño experimental: ejemplo

Participant 1

Trial 1: Color Picker A, target color:

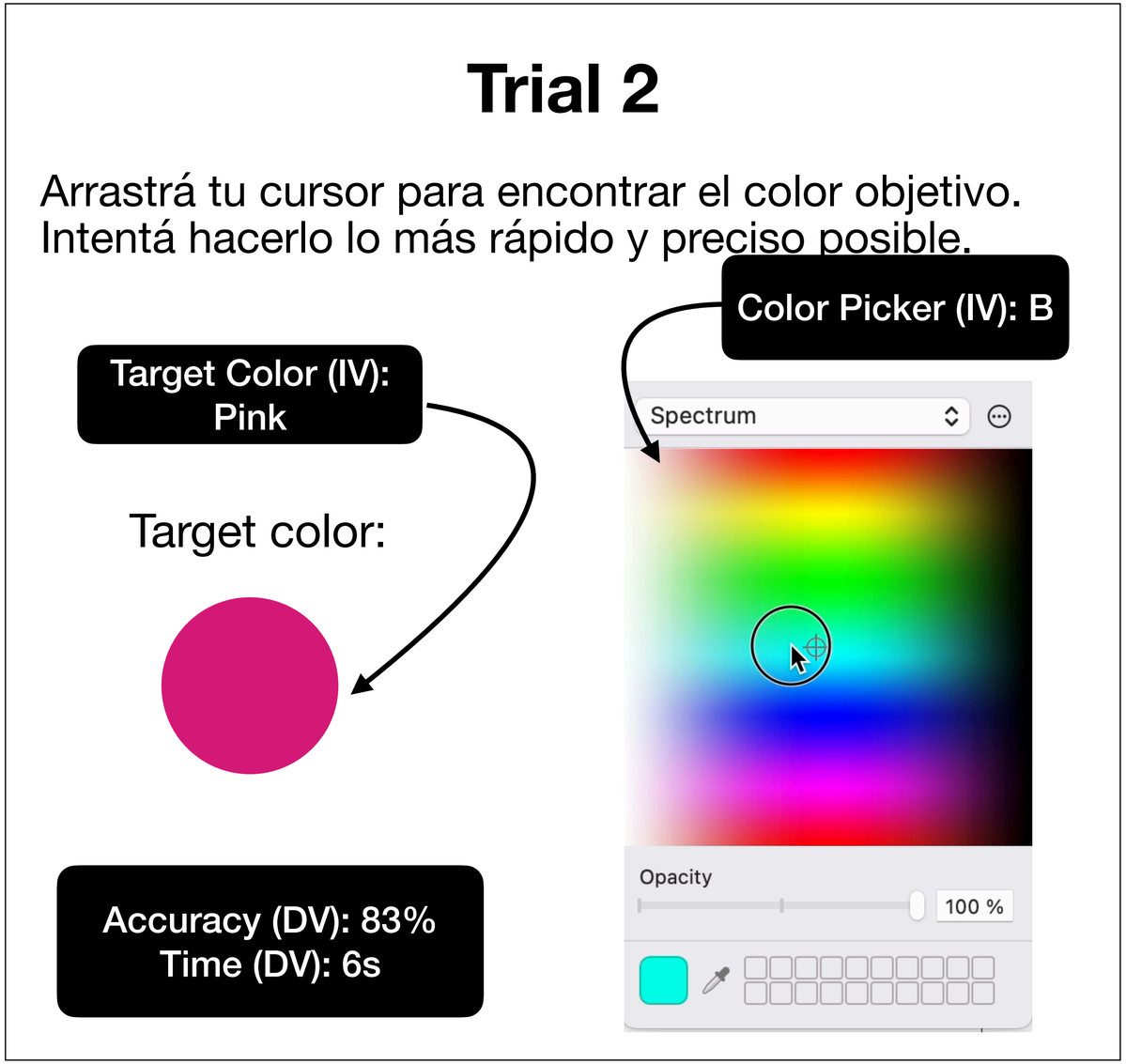


Diseño experimental: ejemplo

Participant 1

Trial 1: Color Picker A, target color:

Trial 2: Color Picker B, target color:



Diseño experimental: ejemplo

Participant 1

Trial 1: Color Picker A, target color:

Trial 2: Color Picker B, target color:

Trial 3: Color Picker A, target color:

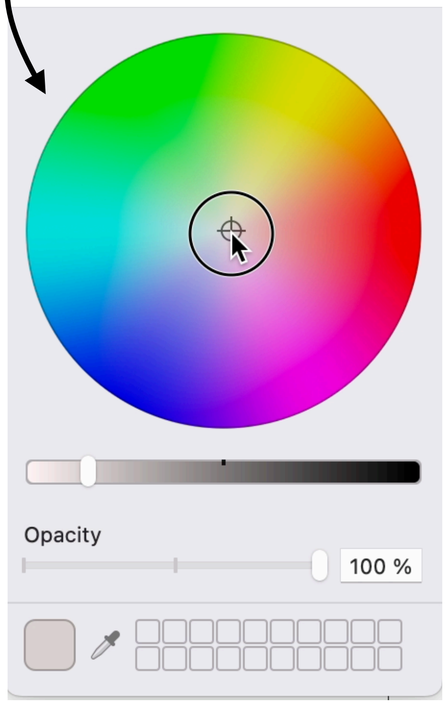
Trial 3

Arrastrá tu cursor para encontrar el color objetivo.
Intentá hacerlo lo más rápido y preciso posible.

Color Picker (IV): A

Target Color (IV): Cyan

Target color:



Accuracy (DV): 91%
Time (DV): 7s

Diseño experimental: ejemplo

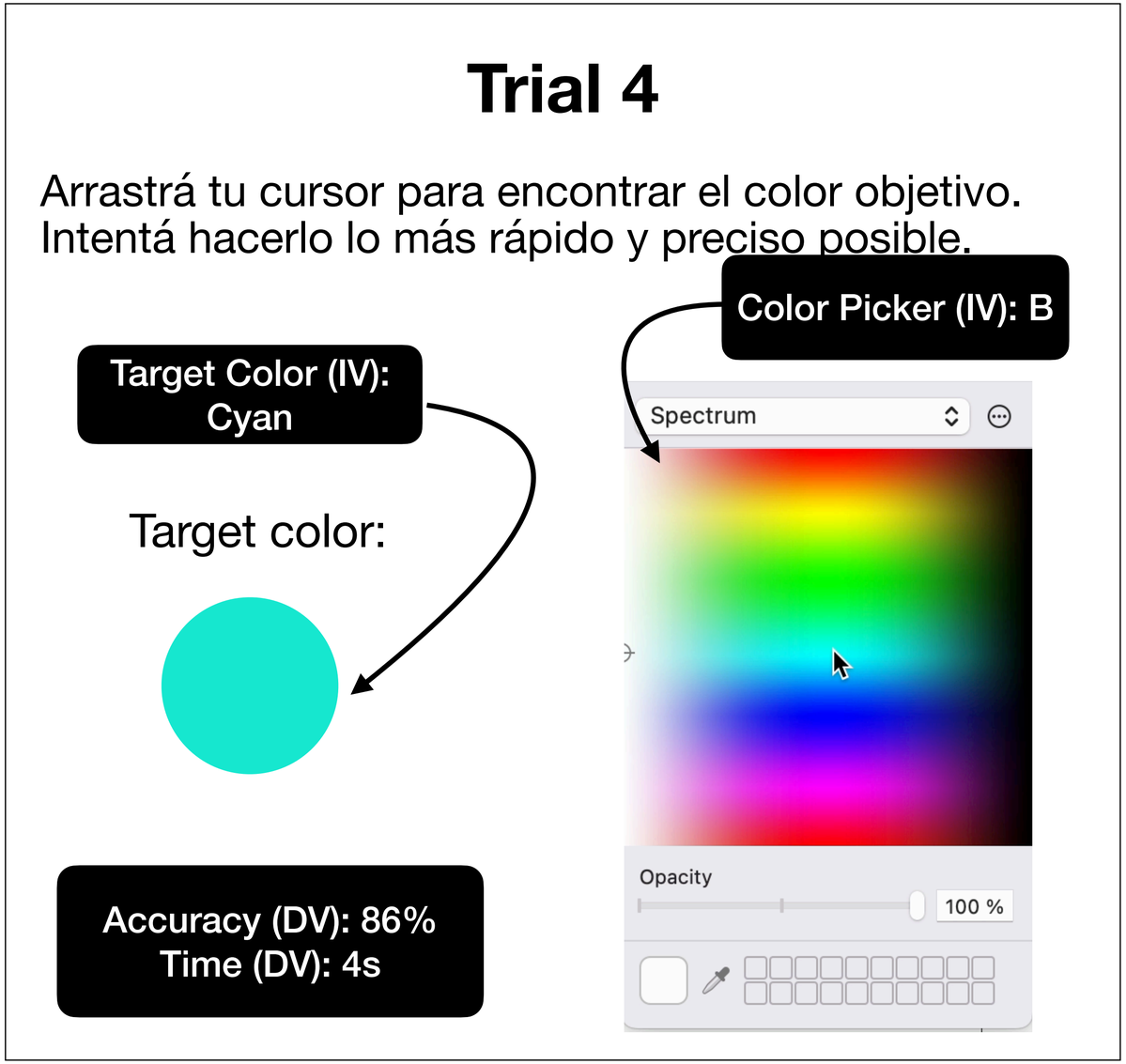
Participant 1

Trial 1: Color Picker A, target color:

Trial 2: Color Picker B, target color:

Trial 3: Color Picker A, target color:

Trial 4: Color Picker B, target color:



Diseño experimental: ejemplo

Participant 1

Trial 1: Color Picker A, target color:

Trial 2: Color Picker B, target color:

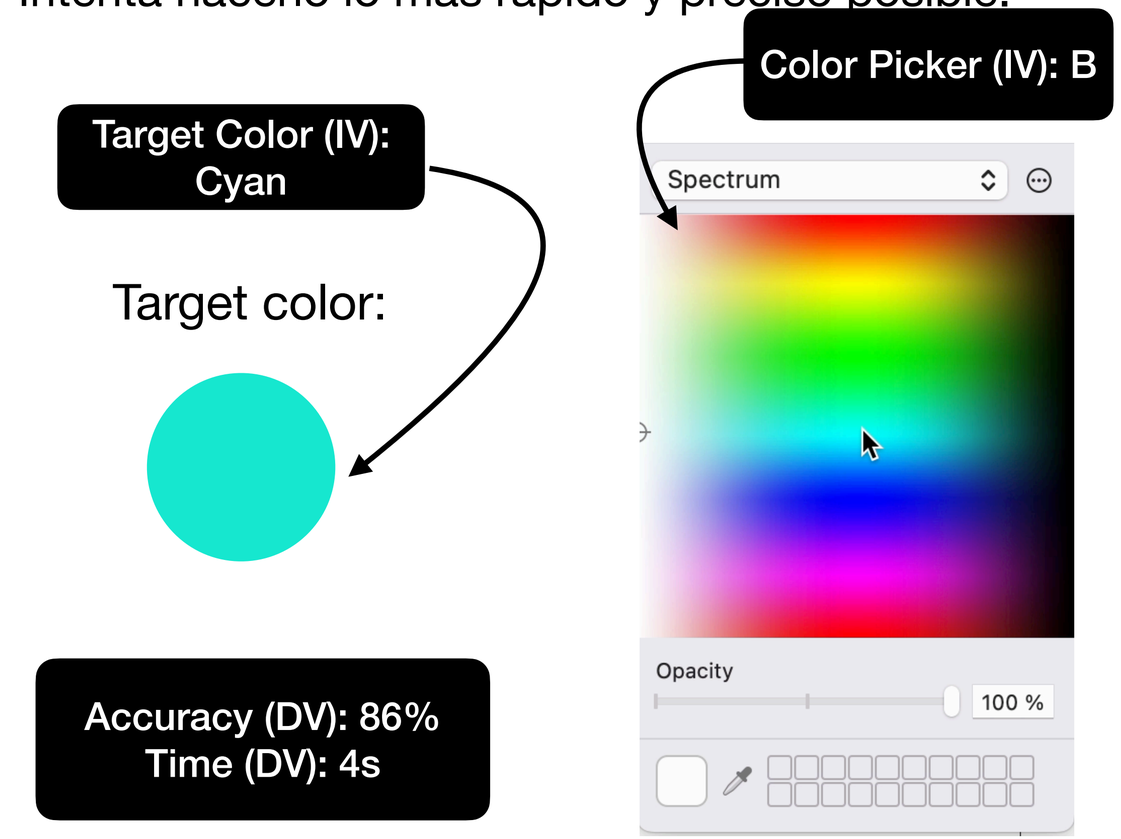
Trial 3: Color Picker A, target color:

Trial 4: Color Picker B, target color:

Should Participant 2 do the same trials in the same order?

Trial 4

Arrastrá tu cursor para encontrar el color objetivo.
Intentá hacerlo lo más rápido y preciso posible.



Diseño experimental: ejemplo

Participant 1

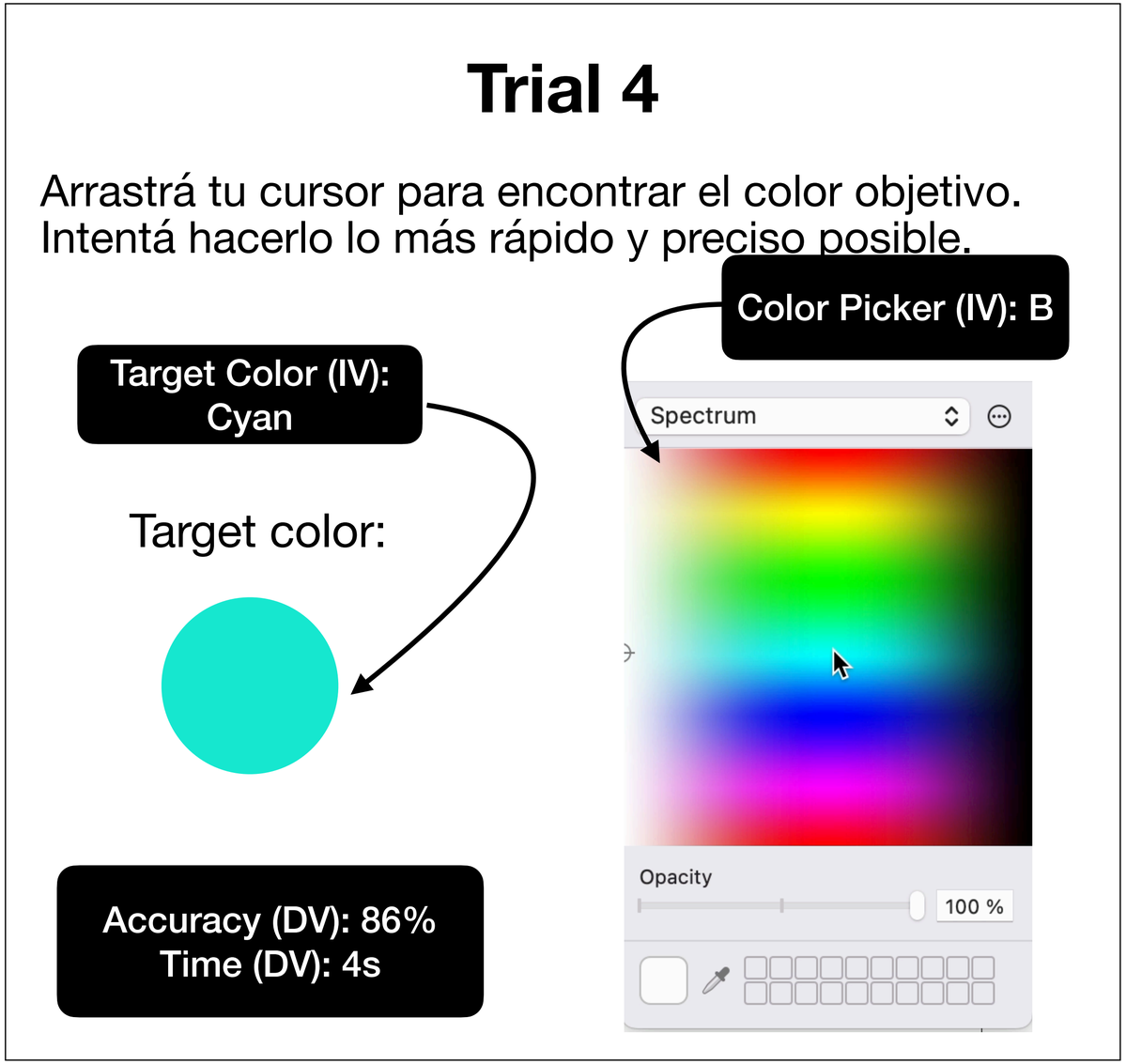
Trial 1: Color Picker A, target color:

Trial 2: Color Picker B, target color:

Trial 3: Color Picker A, target color:

Trial 4: Color Picker B, target color:

Si ahora viene el Participante 2: debería hacer el experimento exactamente como el Participante 1?



Preguntas hasta acá?

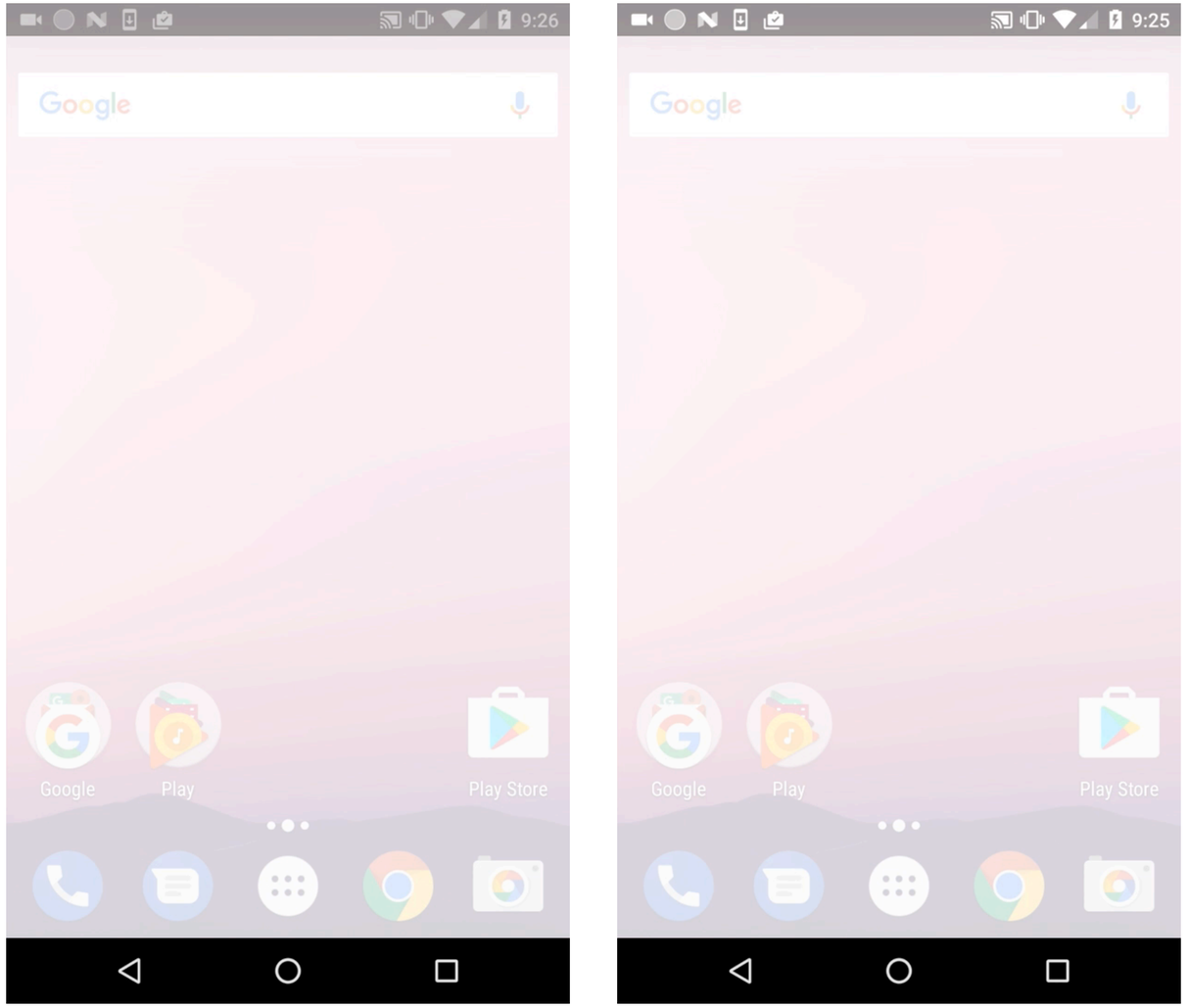
Ejemplo basado en un experimento publicado

Fieldward and Pathward:
Dynamic Guides for Defining Your Own Gestures

Joseph Malloch Carla F. Griggio Joanna McGrenere Wendy E. Mackay

CHI 2017

What if users could create their own gestures?



Finger app for Android

User-defined gestures are easier to remember than pre-existing gestures

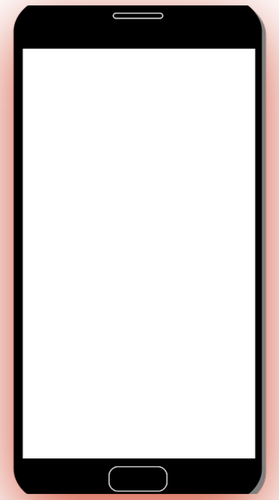
Nacenta et. al (CHI '13)

Users ignore how the recognizer works,
and often create gestures too similar to others

Long et. al (CHI ’99)

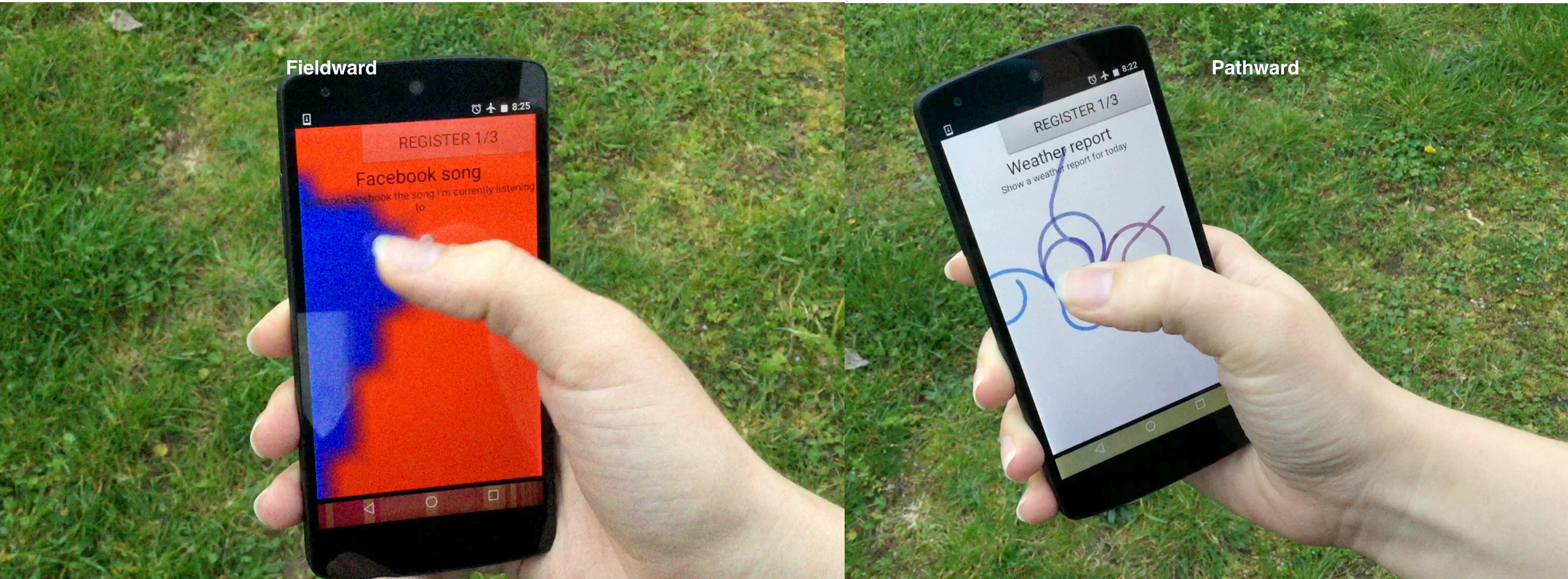
Easy to remember

Easy to recognize



Fieldward and Pathward:

Two dynamic guides with progressive feedforward that **visualize the available space for new gestures**

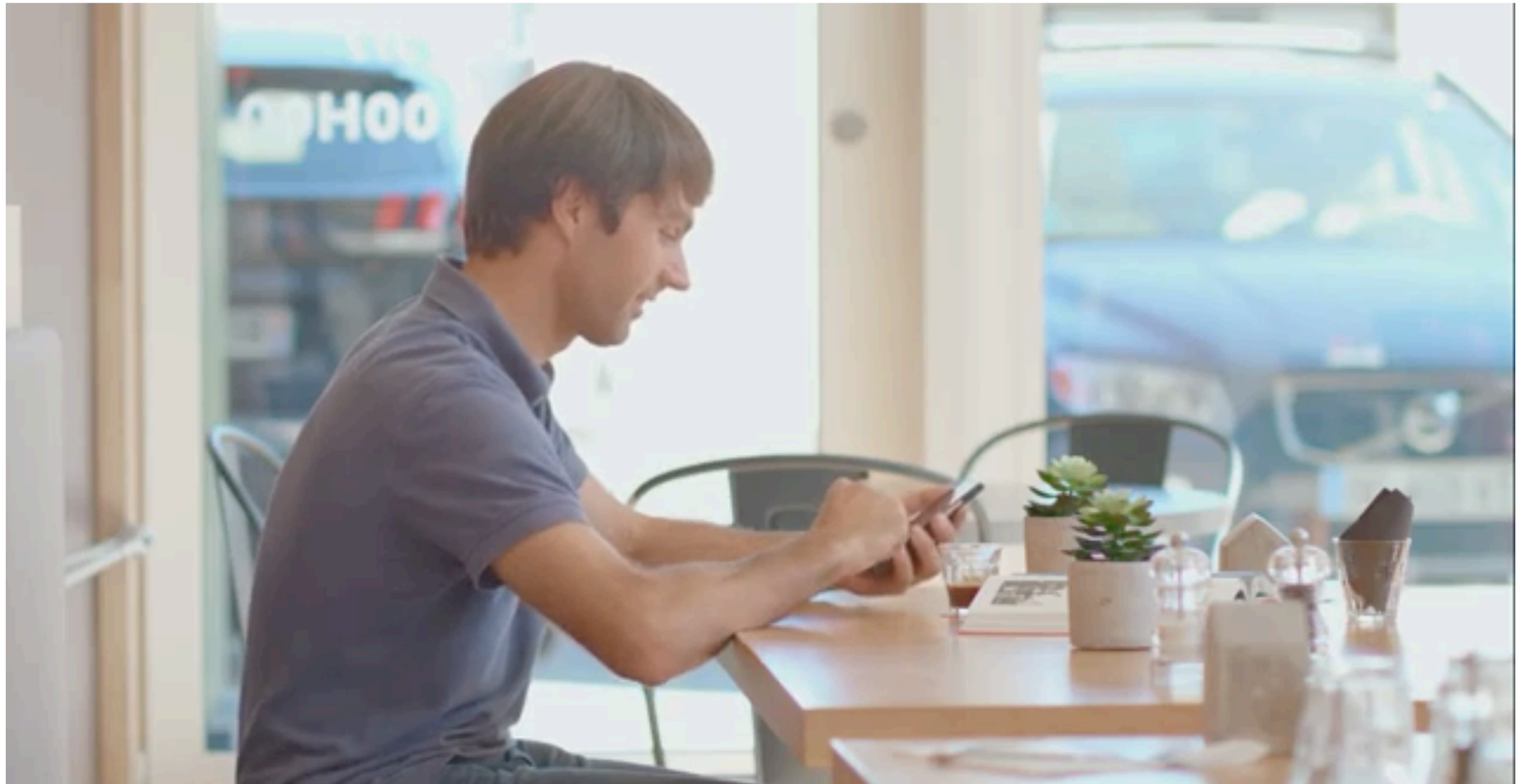


Feedforward

Técnicas para permitirle al usuario anticipar cuál va a ser el resultado de una acción.

Feedforward

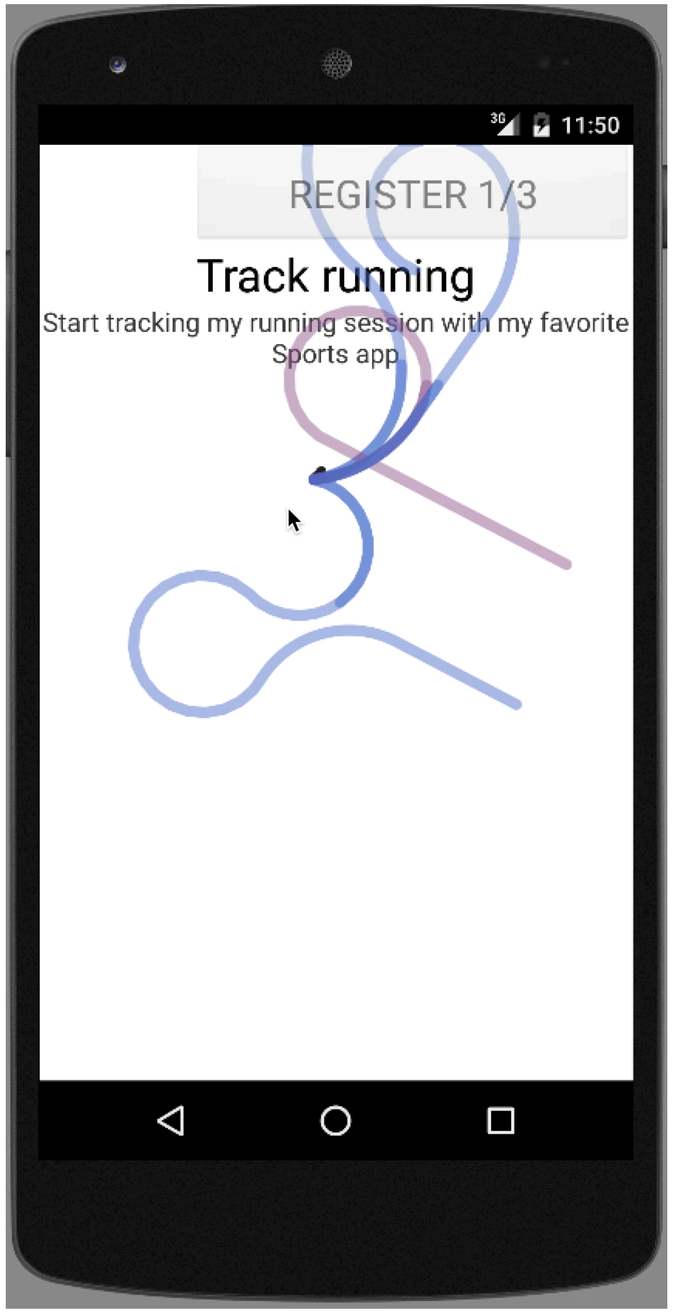
Técnicas para permitirle al usuario anticipar cuál va a ser el resultado de una acción.



Octopocus

Pathward

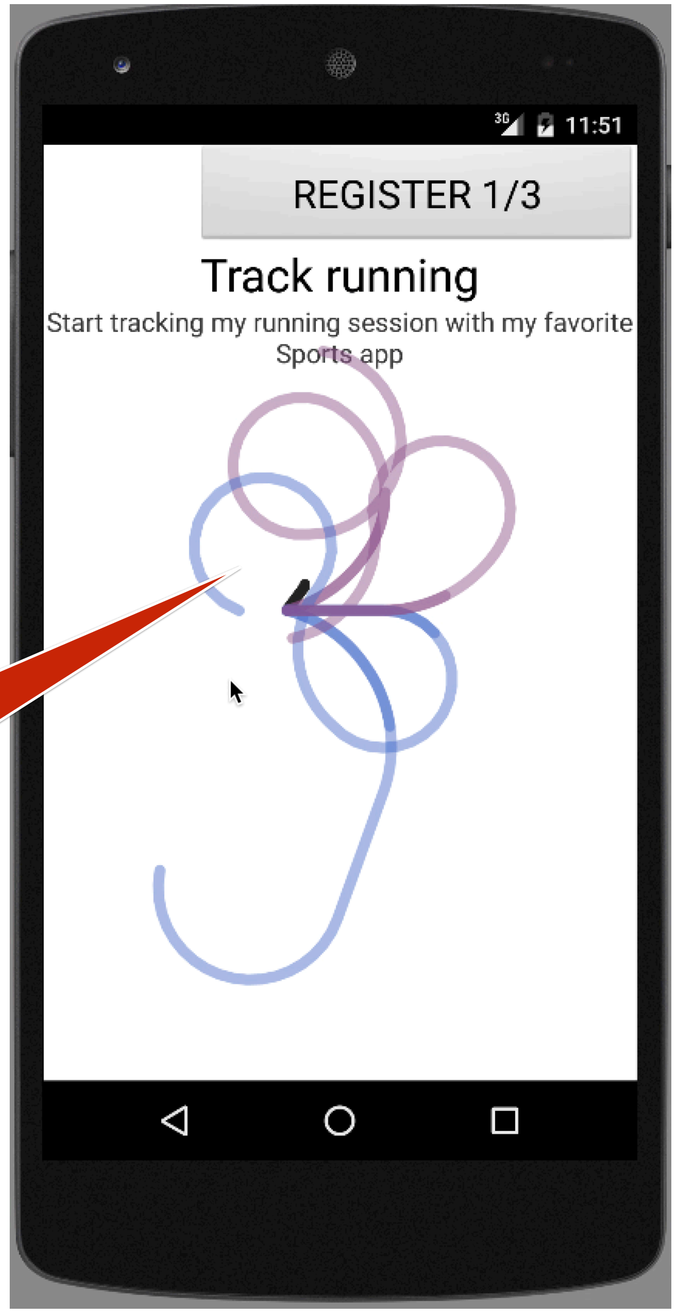
Shows 4 candidate paths
to complete a gesture



Pathward

Shows 4 candidate paths
to complete a gesture

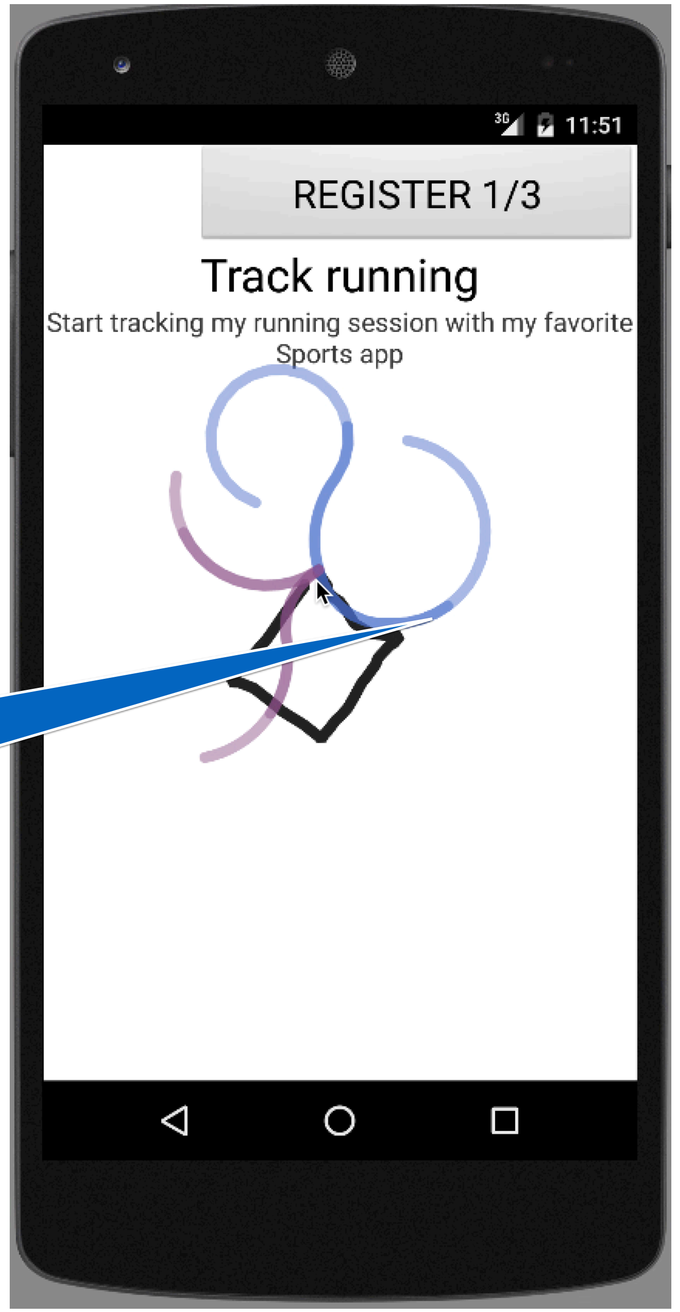
Red paths are gestures that collide with existing gestures



Pathward

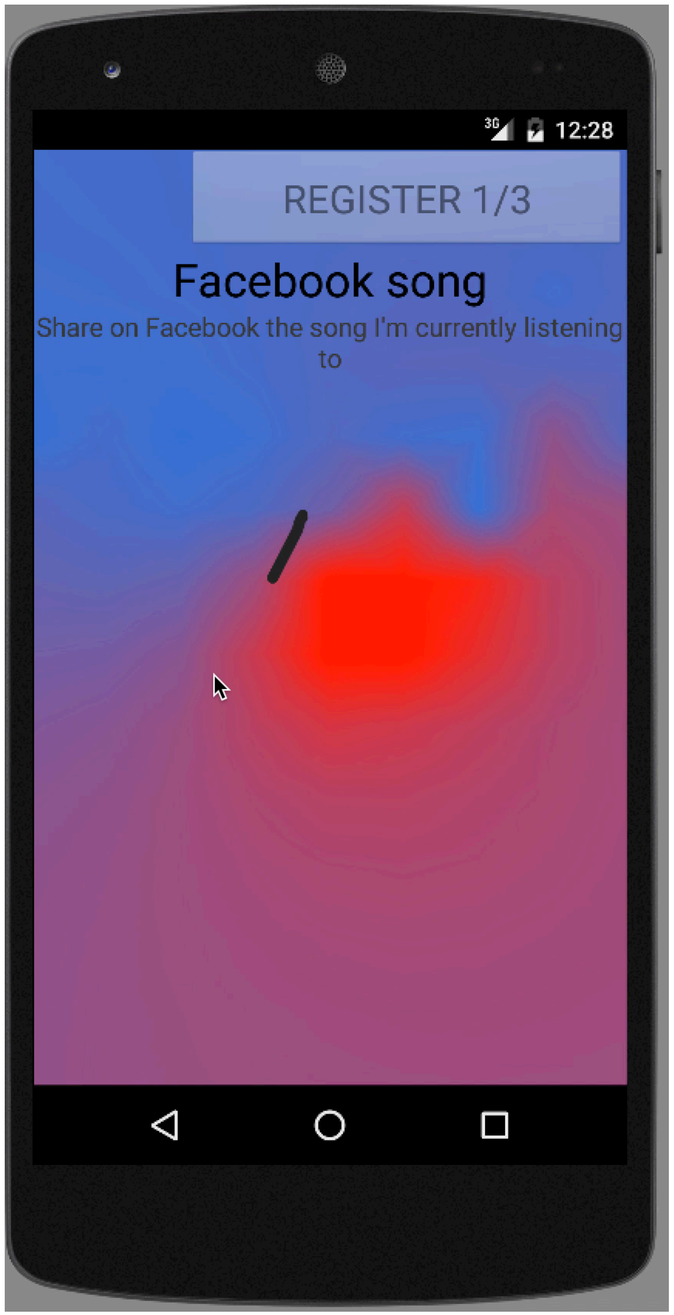
Shows 4 candidate paths
to complete a gesture

Blue paths represent unused recognisable gestures



Fieldward

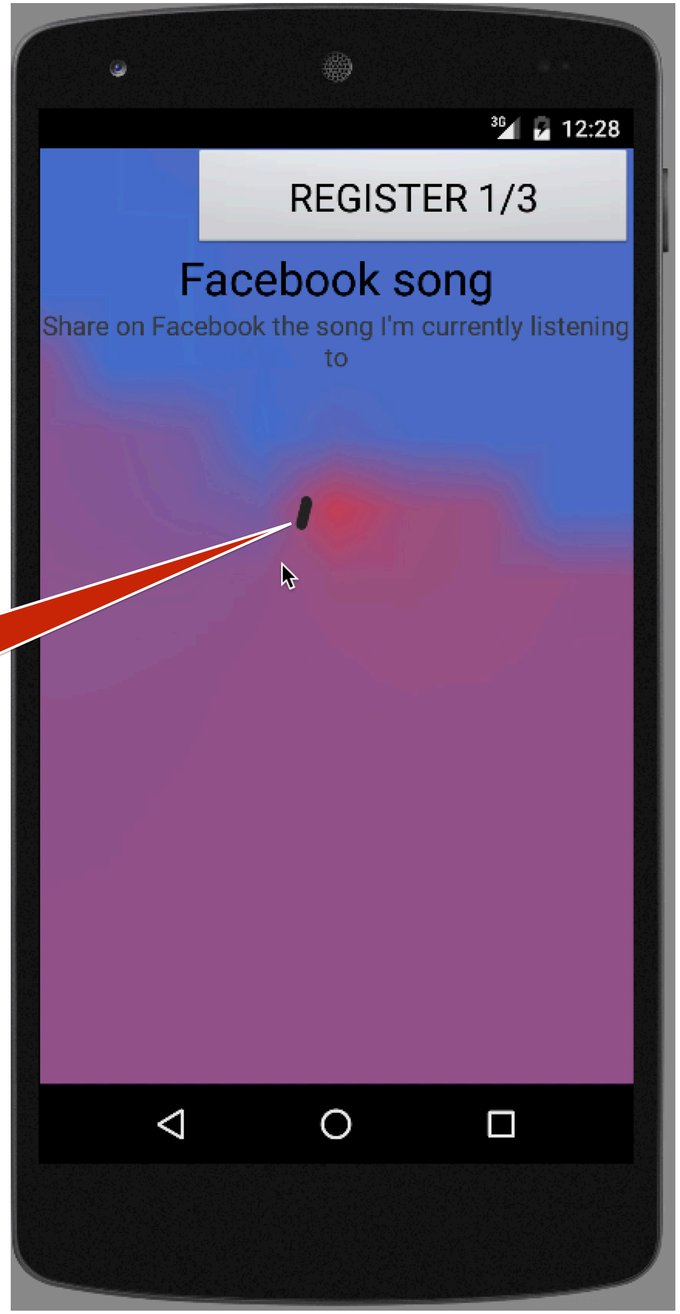
Shows a color gradient
indicating optimal directions
to make a recognizable gesture



Fieldward

Shows a color gradient
indicating optimal directions
to make a recognizable gesture

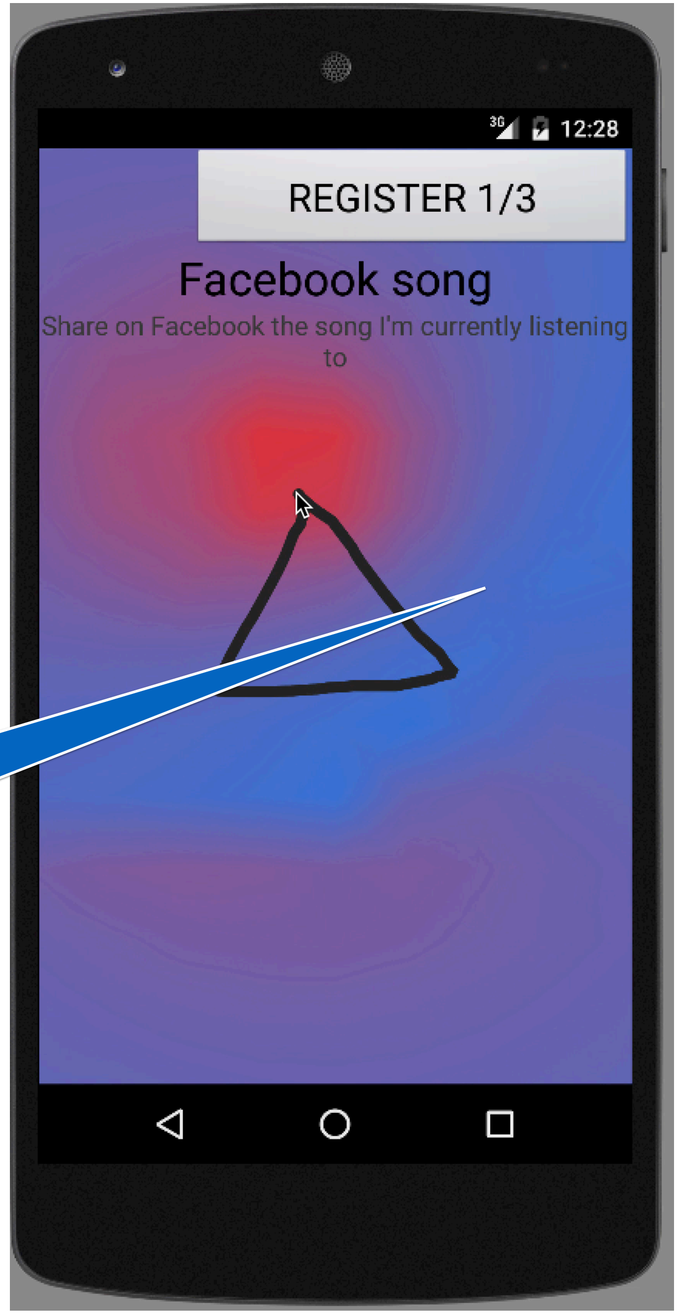
Gestures that end on a **Red** zone collide with existing gestures



Fieldward

Shows a color gradient
indicating optimal directions
to make a recognizable gesture

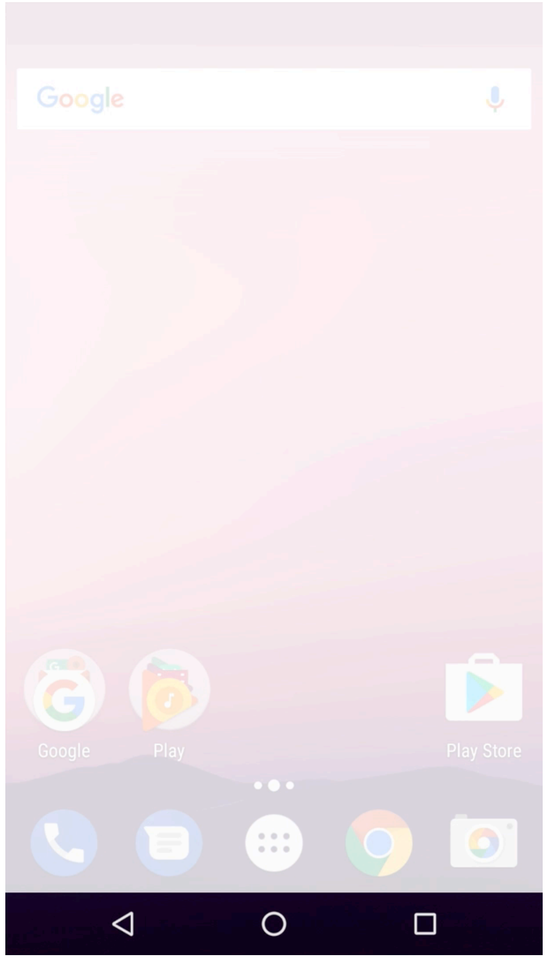
Gestures that end on a **Blue** zone are recognizable



Experiment

Experiment Design

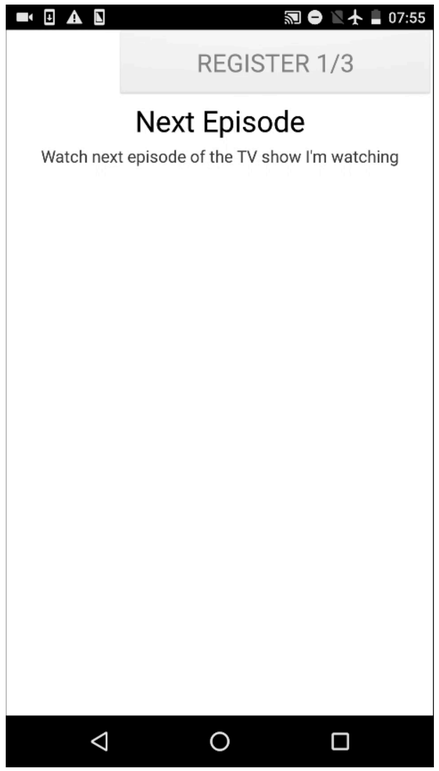
Tarea: crear un shortcut gestual propio



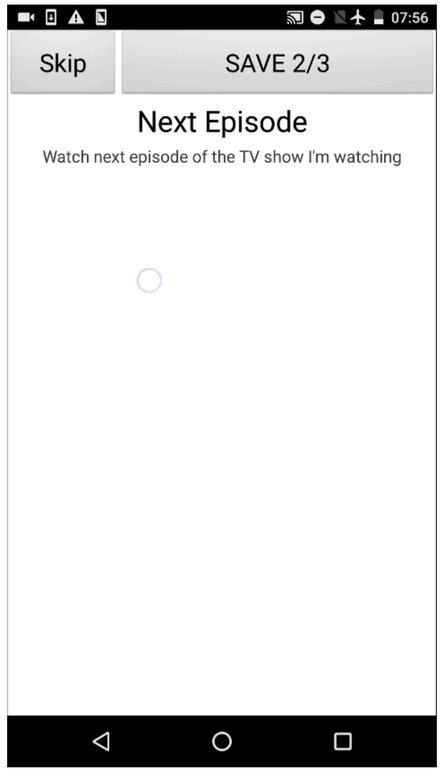
Call partner

Experiment Design

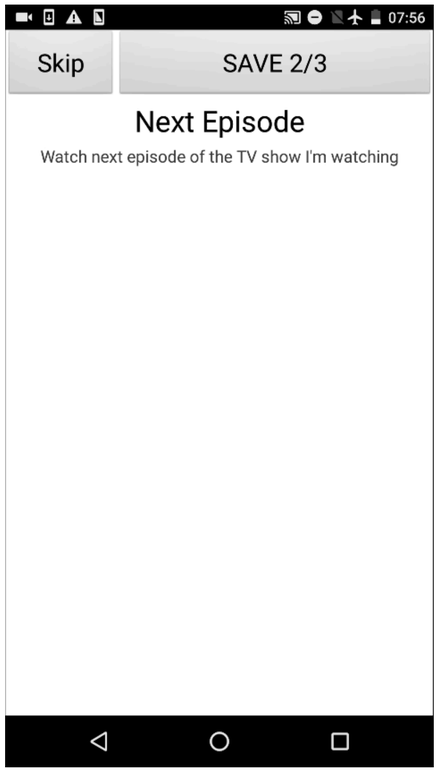
Trial: Registerar un gesto nuevo



Generación
de un nuevo gesto



2do ejemplo

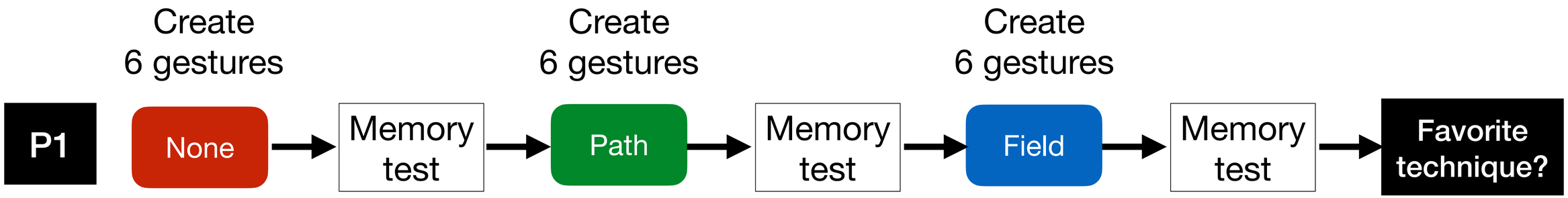


3er ejemplo

Hypothesis: Crear un shortcut gestualidad sin ayuda (None) requiere **más intentos** que con cualquier tipo de ayuda (Field / Path)

Variable Dependiente: cantidad de intentos fallidos (cuántas veces el participante tiene que repetir la tarea de generar un gesto hasta que encuentran uno nuevo que pueden recordar)

Variable Independiente: técnica de feedforward (Field / Path / None)

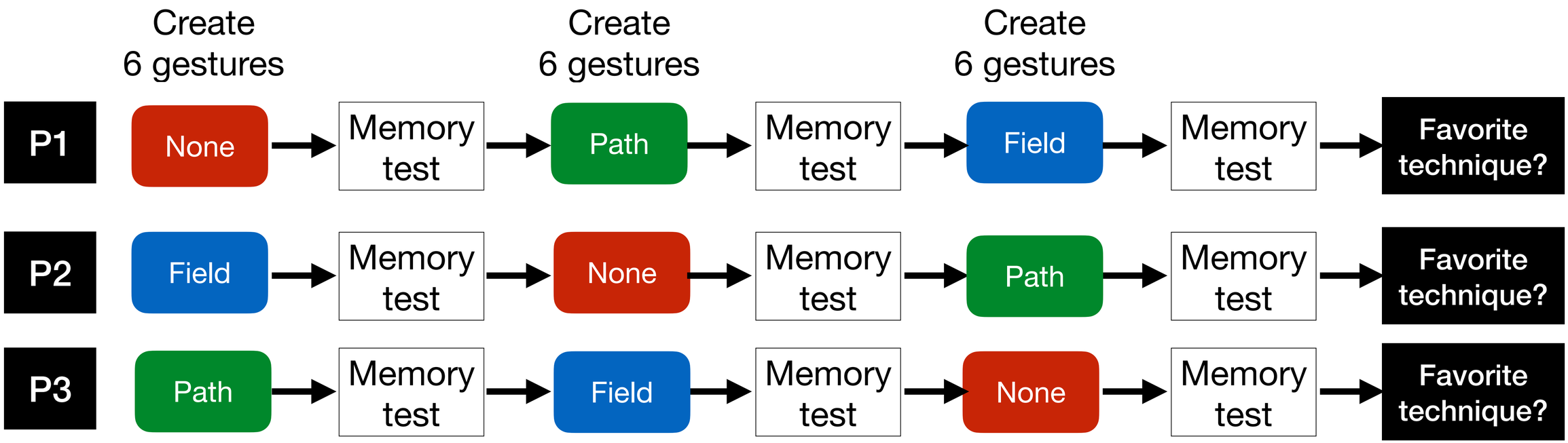


Hay efectos de orden?

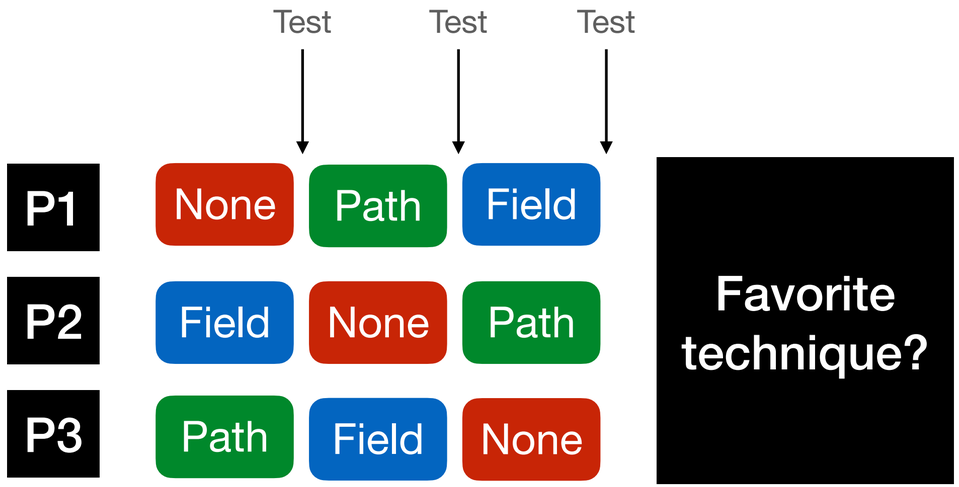
Hypothesis: Crear un shortcut gestualidad sin ayuda (None) requiere **más intentos** que con cualquier tipo de ayuda (Field / Path)

Variable Dependiente: cantidad de intentos fallidos (cuántas veces el participante tiene que repetir la tarea de generar un gesto hasta que encuentran uno nuevo que pueden recordar)

Variable Independiente: técnica de feedforward (Field / Path / None)



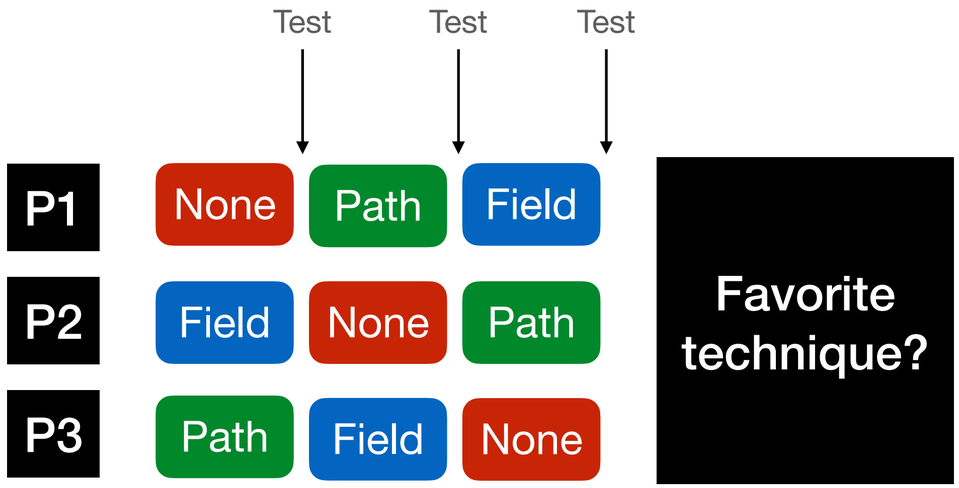
Estudios piloto



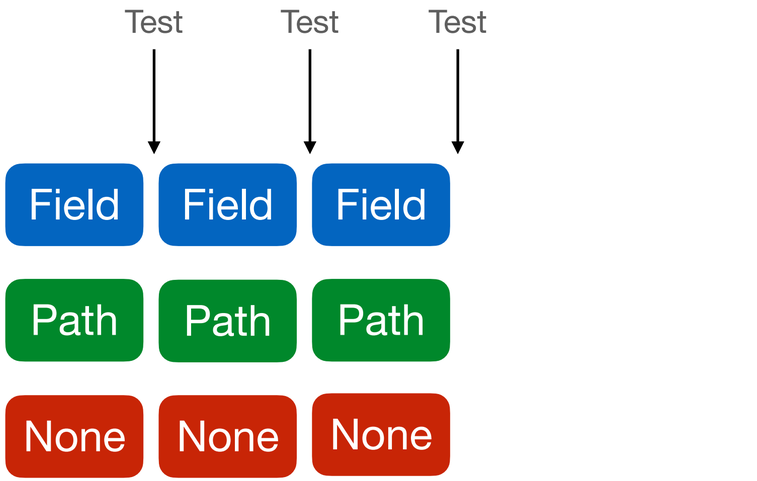
"I represented a Traffic mess by doing a lot of circles on top of each other. I remembered from using Fieldward that it looked like a feasible option".

Ufa, todavía hay efecto de orden :(

Revising the experiment design

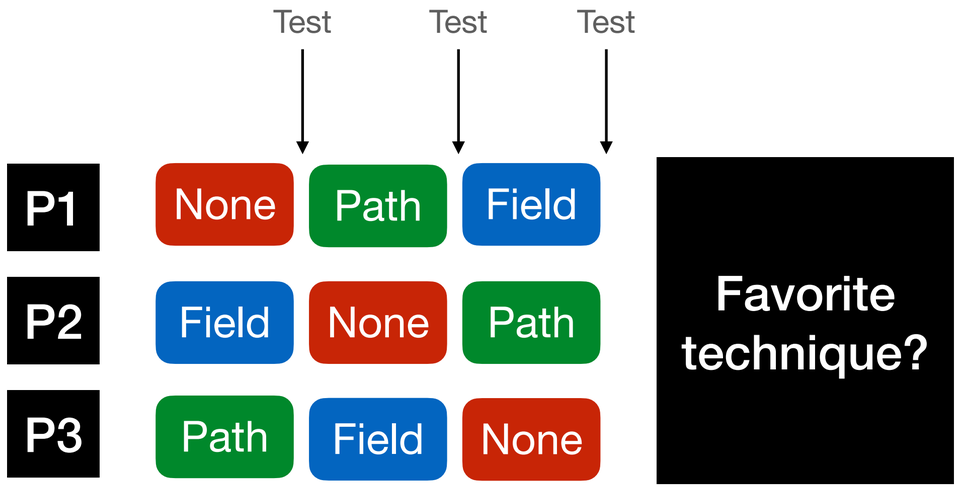


Within-subjects design:
all participants are exposed to all conditions

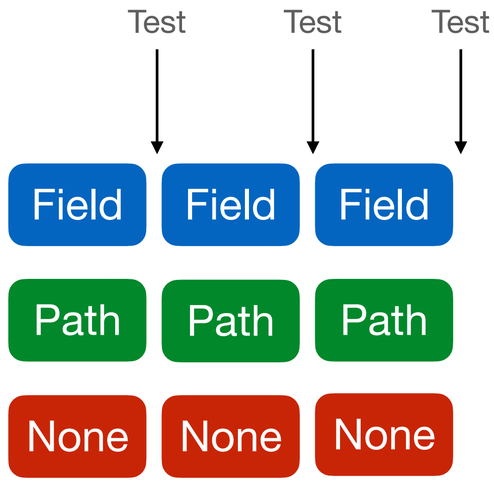


Between-subjects design:
different groups of participants are exposed to different conditions

Revising the experiment design



Favorite technique?



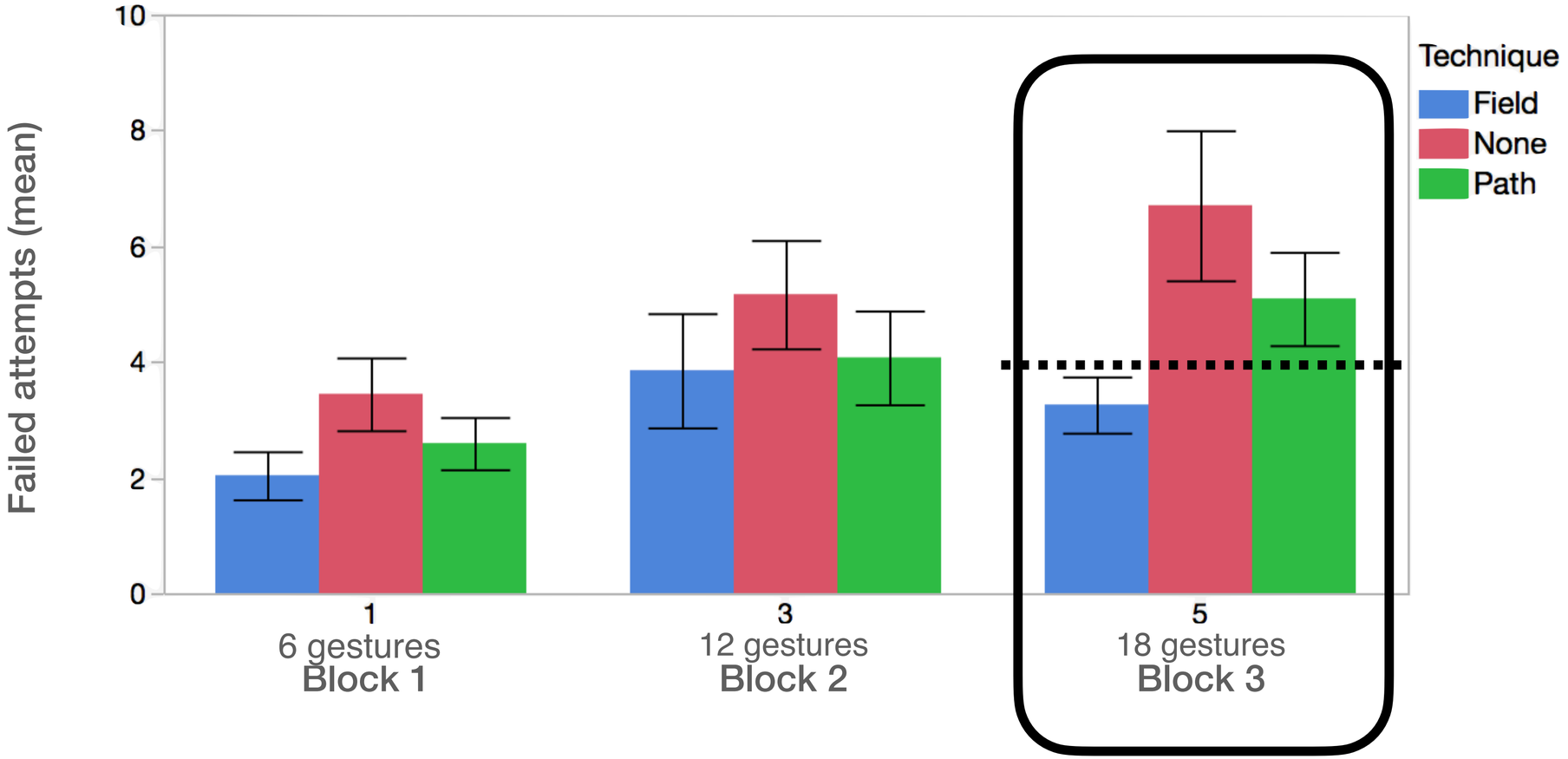
Within-subjects design

Between-subjects design

 # de intentos

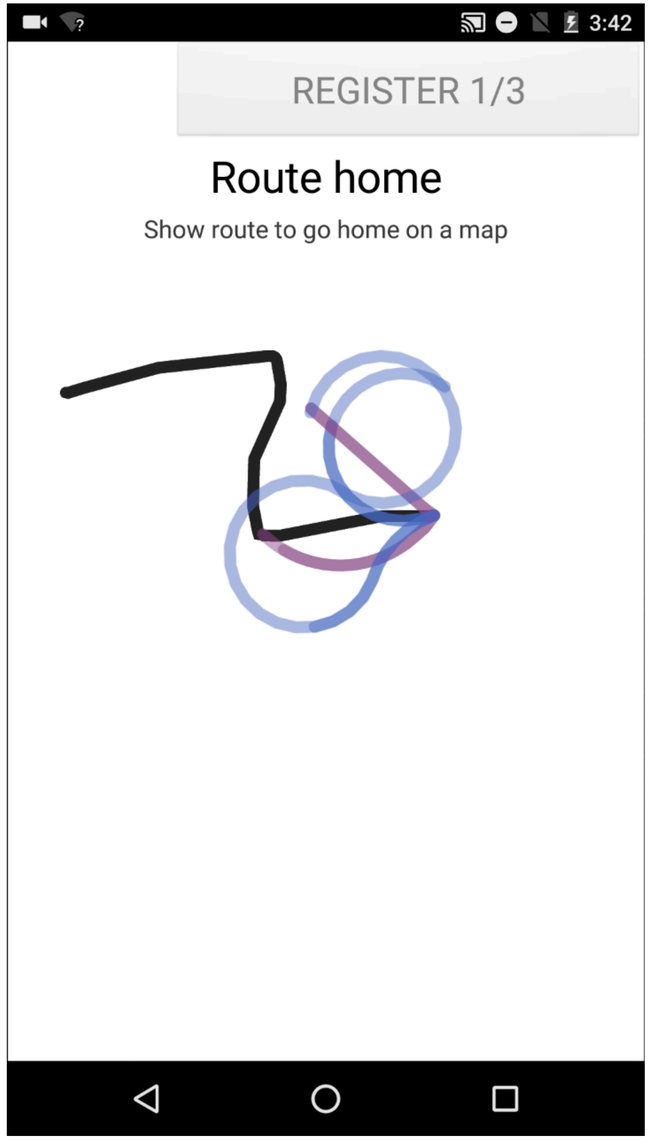
Técnica preferida

Results with 27 participants (9 in each group)

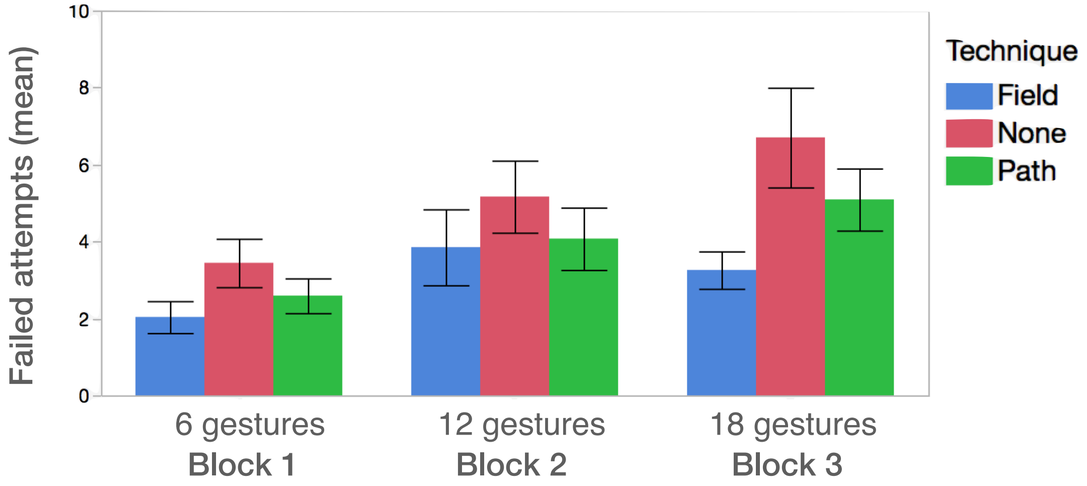


Fieldward gets less failed attempts when the task is the hardest

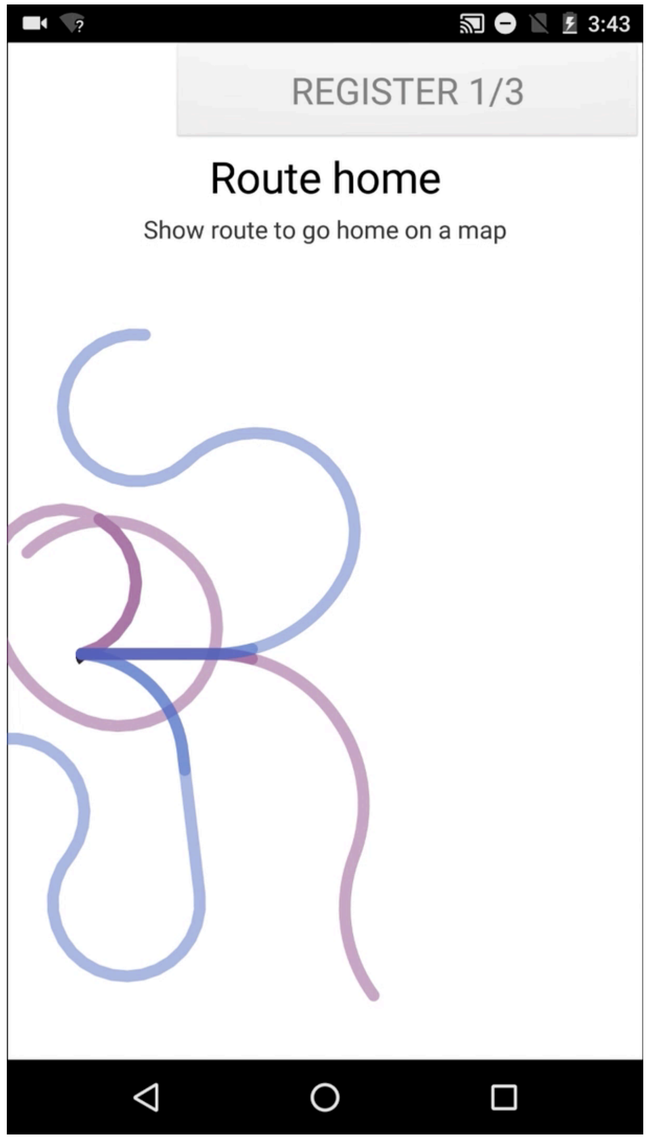
Participants value memorability over recognizability



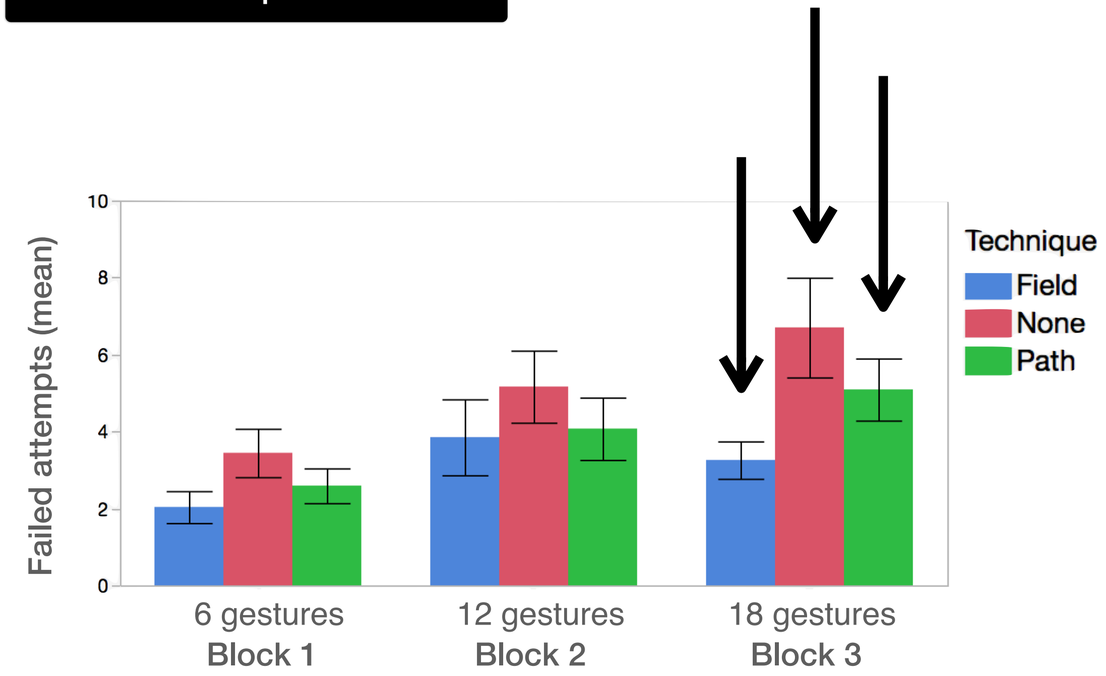
P18 looking for a gesture similar to his own “route home"



Participants value memorability over recognizability

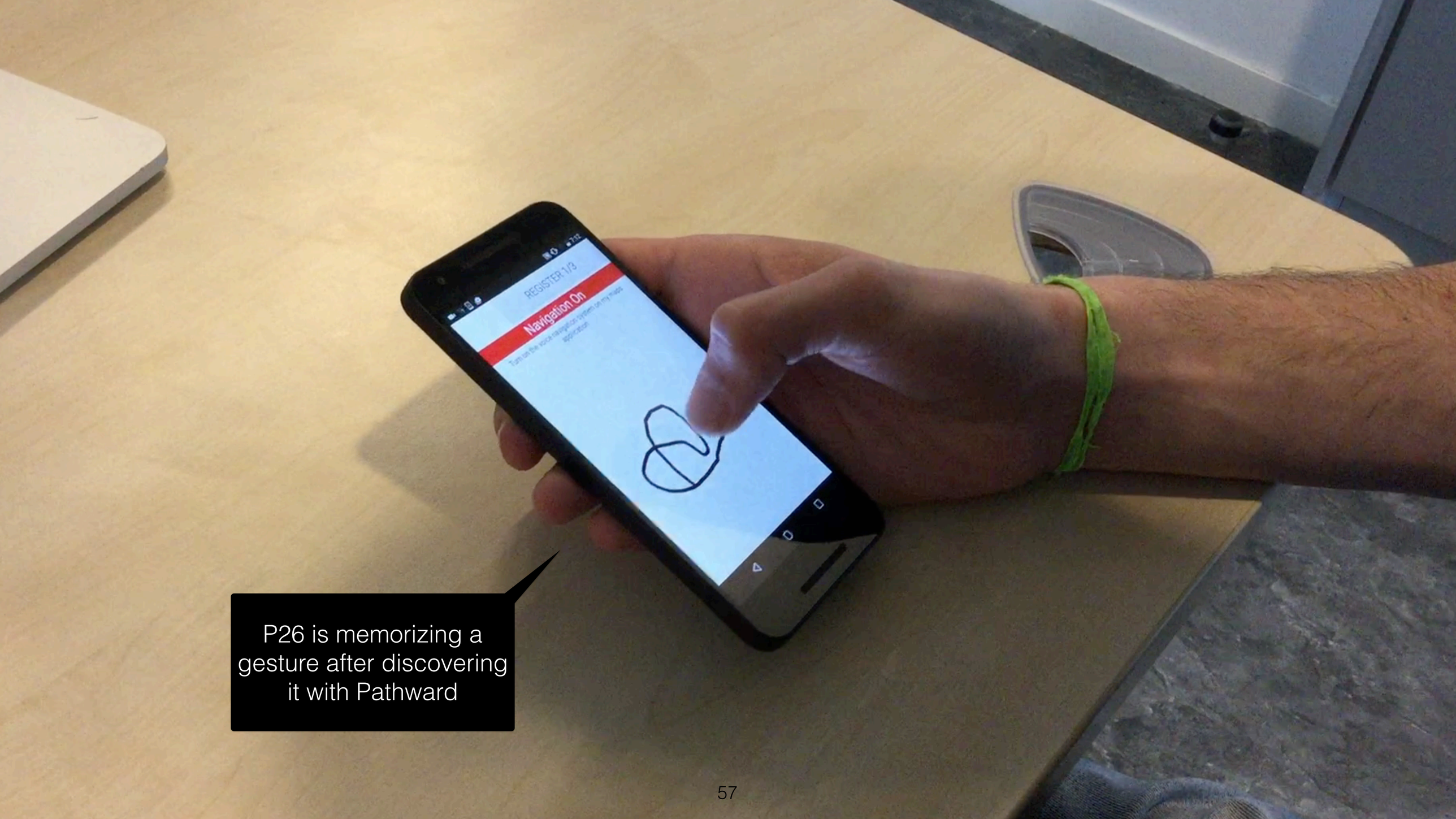


P18 adapting his own gesture by following a blue path



"I get that the lines [in Pathward] wanted to help me, but ... but they help you ***'step by step'****, and I had a* ***complete gesture idea*** *in my mind"*

-P7

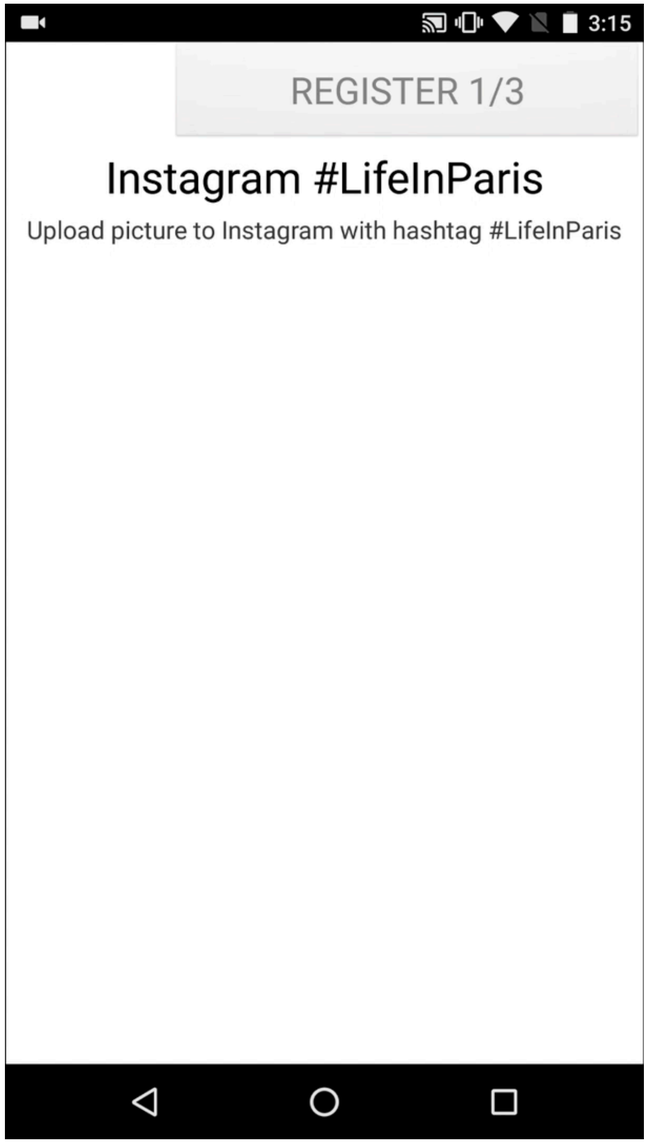


P26 is memorizing a gesture after discovering it with Pathward

"*Fieldward is free enough to help you to create a figure and remember it. (...) It’s hard to follow a path and remember the figure you made*"

-P8

Fieldward better supports user's ideas for memorable gestures



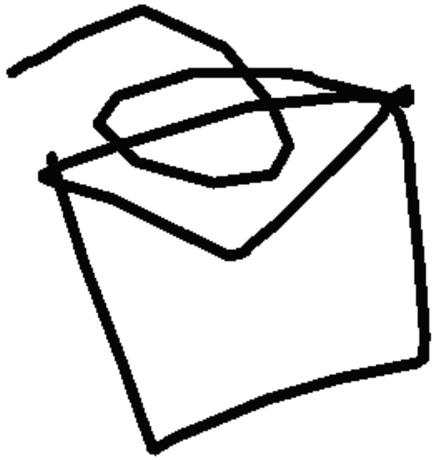
P19 adapting her Eiffel Tower to be recognisable

Question:
“I prefer creating memorable gestures with:"

Preferred Technique

Different strategies adopted to support memorability

Gesture grammars

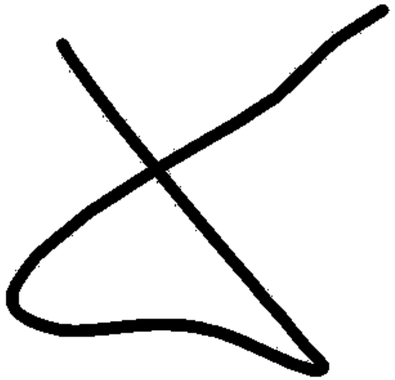


Forward parents:
envelope
+
curl (send)

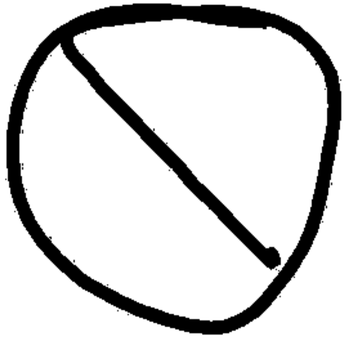
Take selfie:
winky eye (picture)

Different strategies adopted to support memorability

Cultural differences & influences from other devices



Close all:
Cross
(common for
Western users)



Close all:
Crossed circle
(common for
users from Asia)

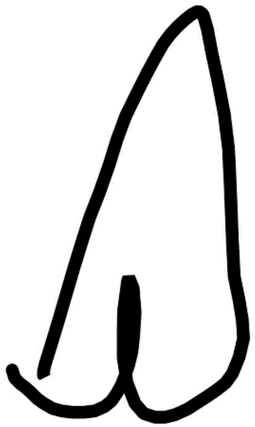
Close all:
Four “swipe right”
gestures
(closing apps in
Android)

Close all:
Command + W
(close window in
MacOS)

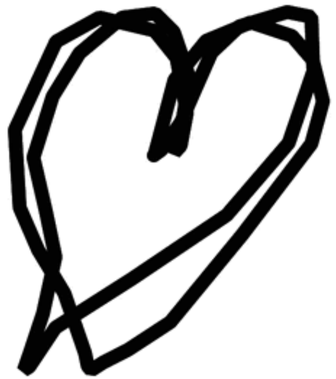
Different strategies adopted to support memorability

Changing the sequence, rotation and length of gestures to preserve its shape

Family Location
Heart



Poke Partner
Rotated heart



Forward parents
Double heart

Experimentos como un método para entender a los usuarios

- El experimento nos permitió evaluar cuál técnica fue más eficiente en ayudarle a los participantes a encontrar gestos nuevos que fueran fáciles de recordar.

- Pero las observaciones que hicimos durante el experimento también nos ayudaron a entender más sobre qué necesitan los usuarios para crear gestos nuevos y cómo piensan

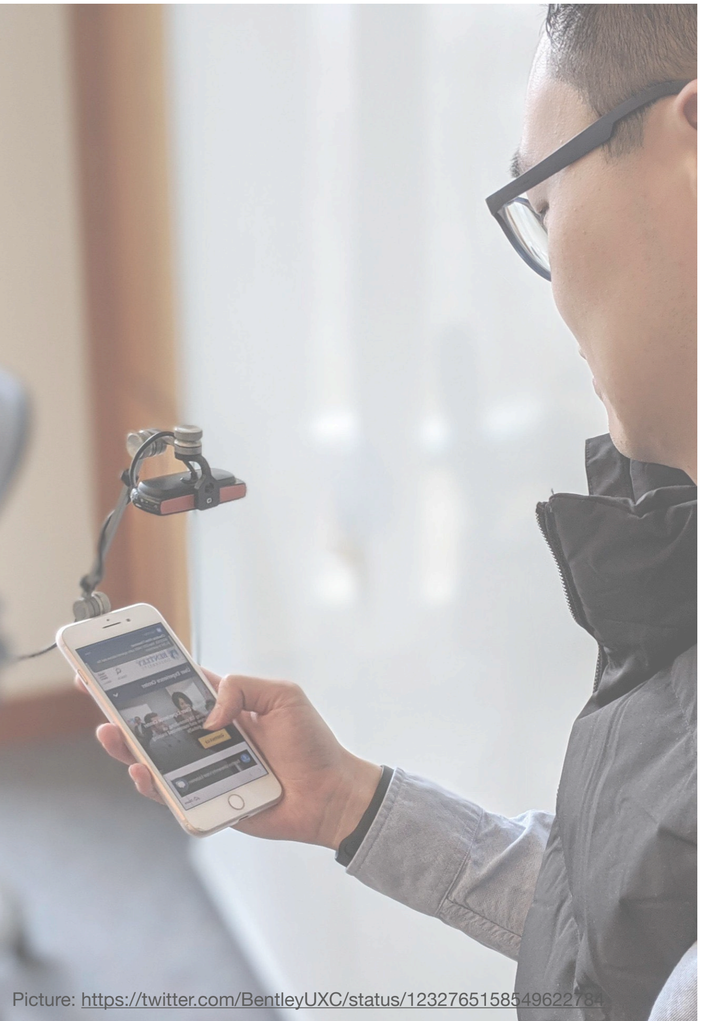
- Entonces, los experimentos no sólo sirven para comparar alternativas de diseño, también pueden ayudar a entender mejor a los usuarios por fuera de métricas específicas

Cuasi-experimentos

- Se definen tareas y se controlan algunas variables, pero sin el objetivo de encontrar una relación causa-efecto entre los diseños a evaluar y métricas de uso.
- Los datos son mayormente cualitativos (e.g., observaciones, o entrevistas después de completar el estudio), y si hay datos cuantitativos son **reportados** (cuestionarios estilo NASA-TLX o con preguntas propias)
- Sirven para entender en detalle y de manera estructurada cómo los usuarios entienden y usan el sistema, encontrar errores inesperados o ventajas inesperadas
- A partir de las observaciones hechas en un cuasi-experimento, se pueden proponer hipótesis para validar en un experimento.

Qué pasa afuera del laboratorio?

Qué pasa afuera del laboratorio?



Picture: https://twitter.com/BentleyUXC/status/1232765158549622784

Lab	Vida real
Controlamos el contexto	El contexto puede variar inesperadamente
El sistema se usa en aislamiento	El sistema se usa en una ecologia de artefactos
Tareas pre-definidas	Acción situada (los planes cambian, y los objetivos se resuelven con lo que hay a mano)

Validez Ecológica (Ecological Validity)

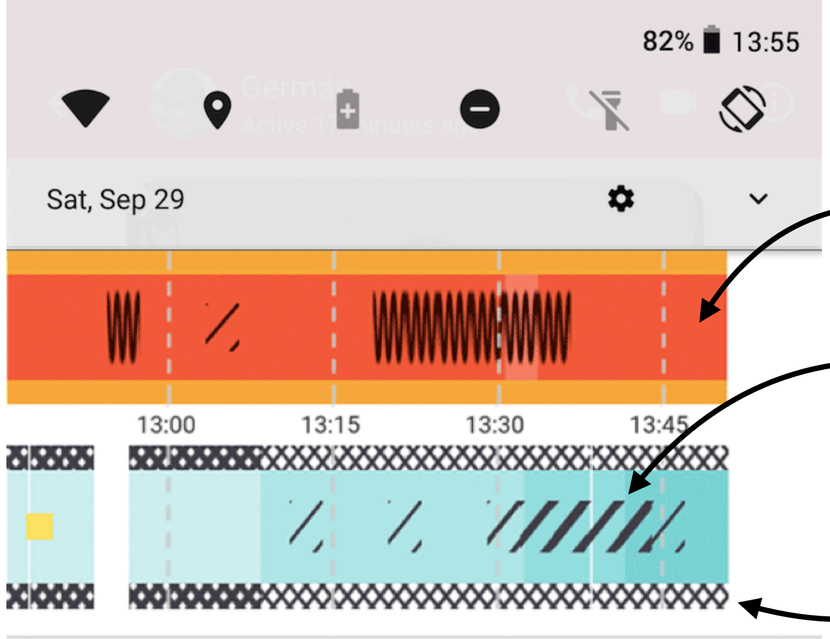
Qué tan probable es que los resultados del estudio en el laboratorio se reflejen en el mundo real?

Estudios de campo

(Field Studies)

Se realizan en “condiciones reales”, en el mundo real

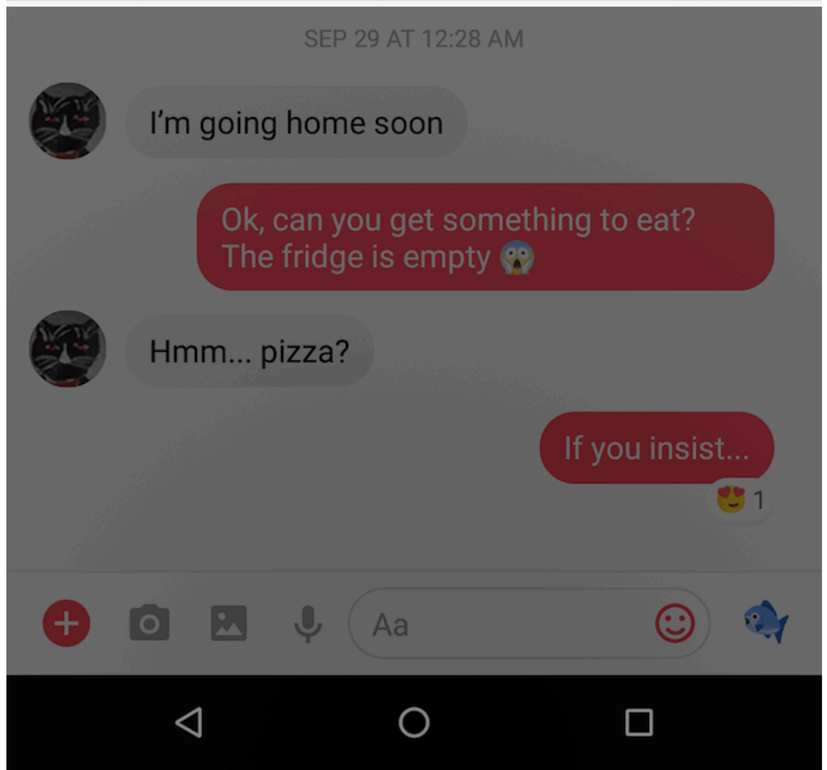
More control	⟷	Less control
Participants use our pre-set devices		Participants use their own devices
Participants are instructed to perform specific tasks		Participants use the system as they see fit
Participants use the system at specific times in the day		Participants use the system whenever they want
Participants use different versions of the system determined by us		Participants choose what version of the system to use according to their needs



Distance from home

Steps

Battery level



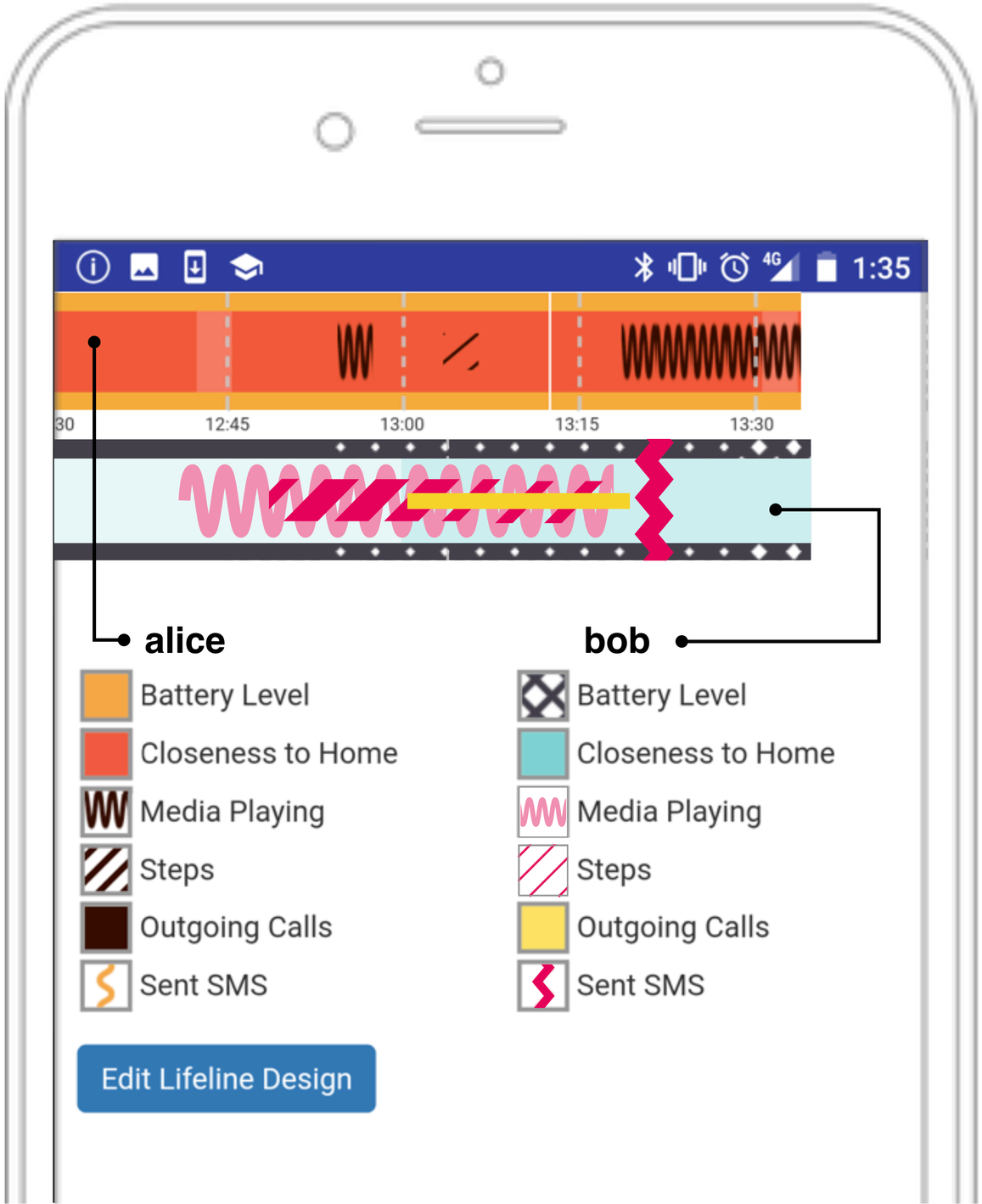
Goal: understand how sharing contextual information affected couple's online communication

One-month field study with 9 couples, using the system on their own smartphones

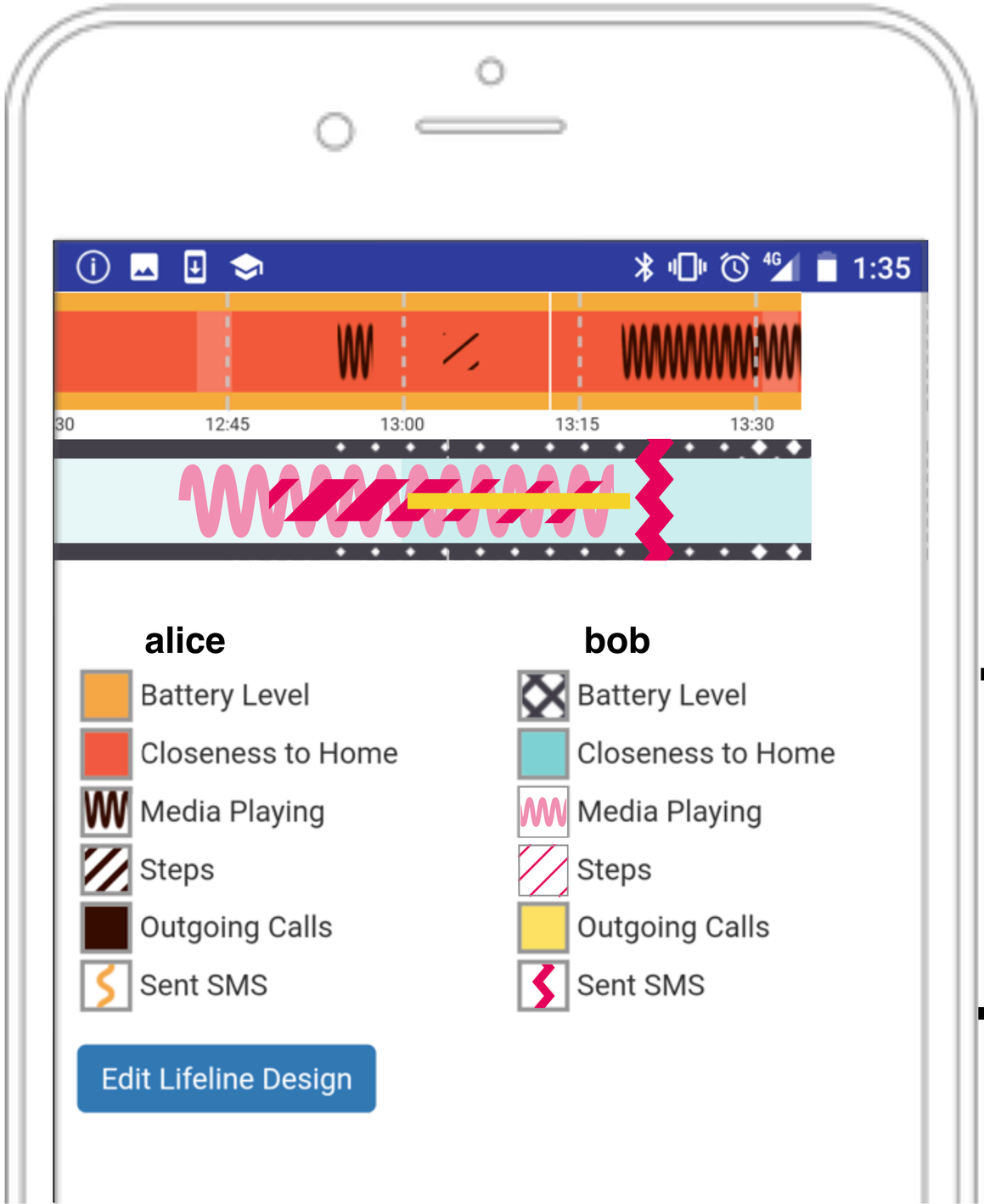
Lifelines

Griggio, Nouwens, McGrenere and Mackay. CHI 2019.
https://doi.org/10.1145/3290605.3300853

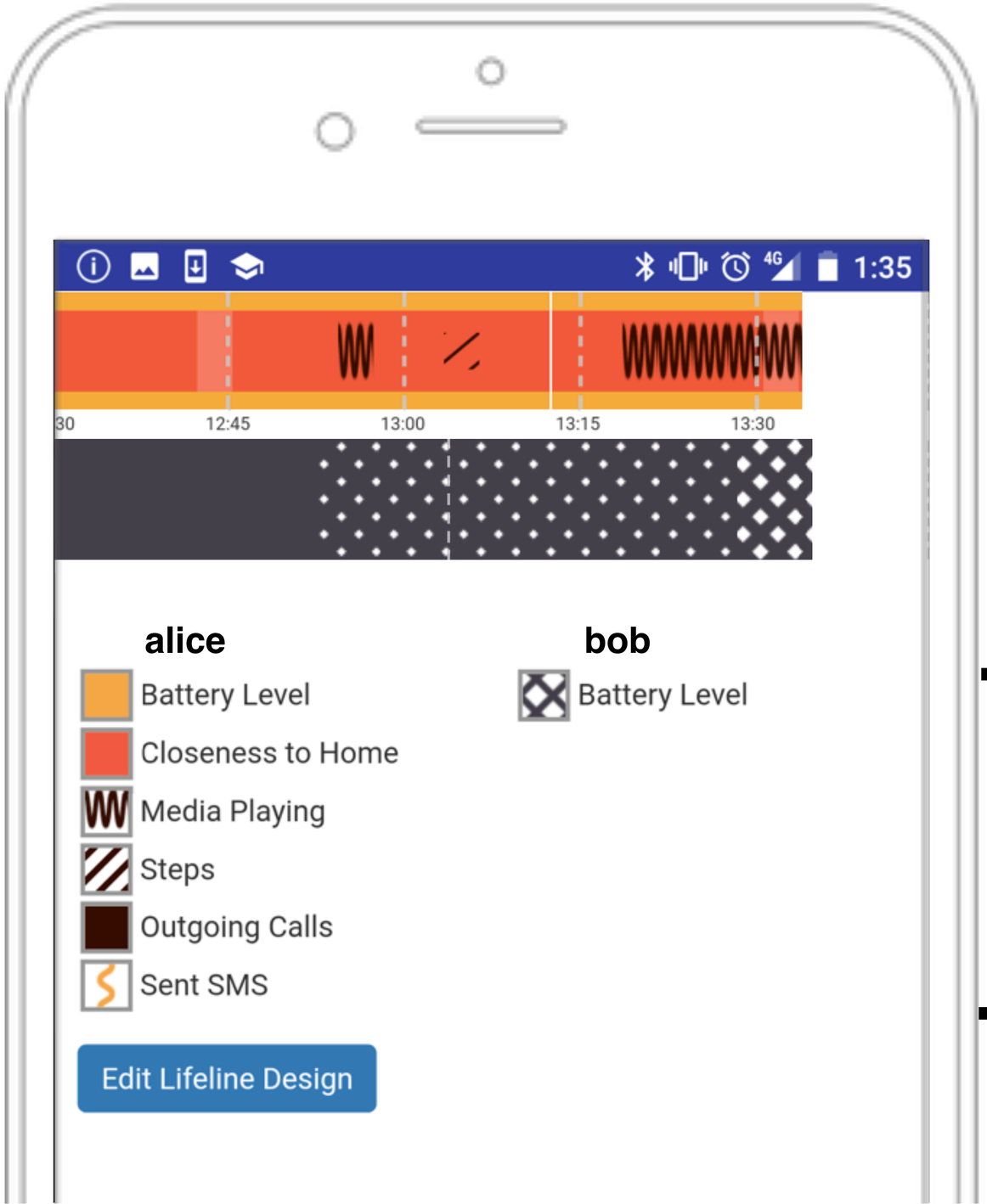
The Lifelines technology probe



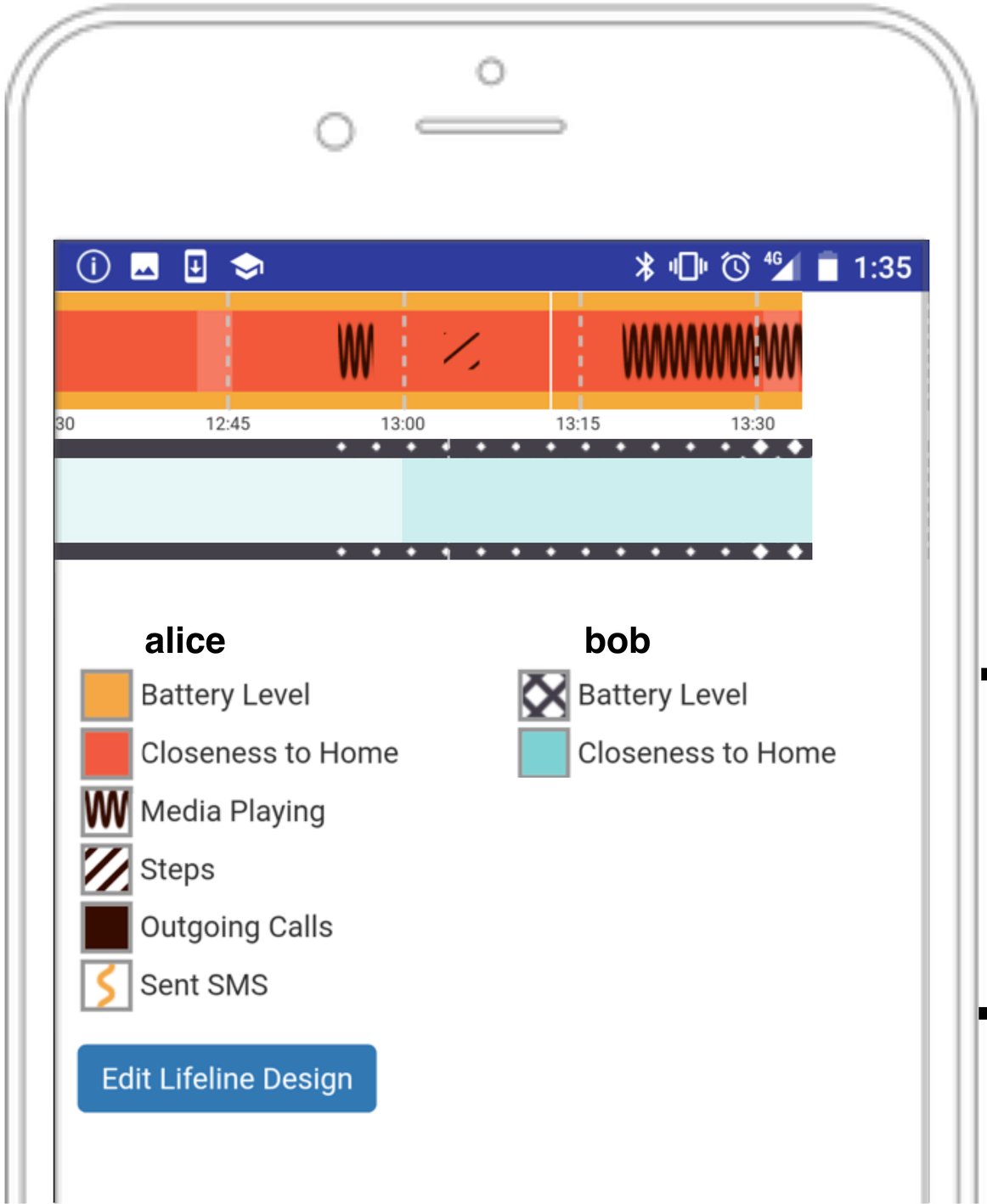
The Linebuilder



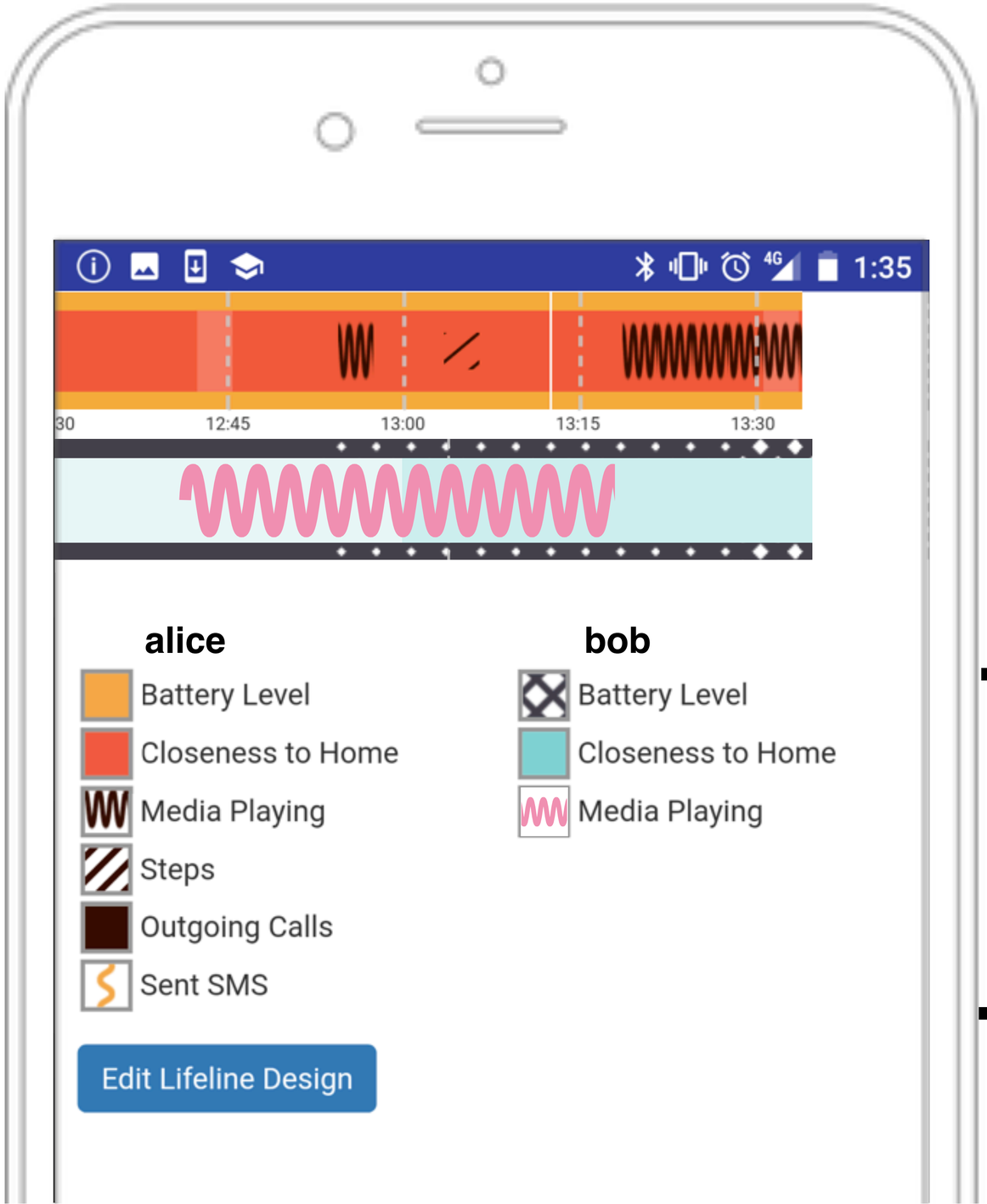
6 streams of contextual information



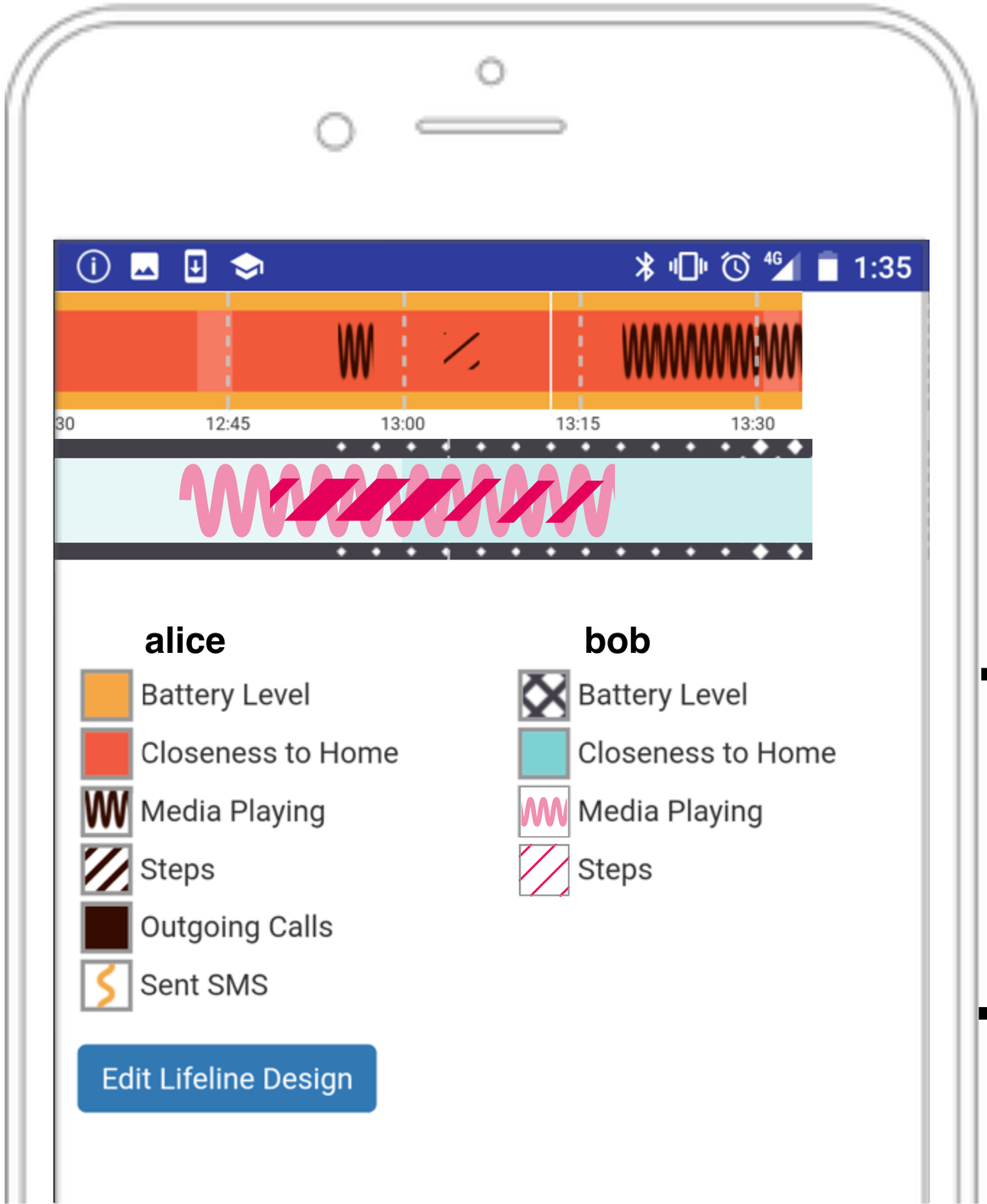
6 streams of contextual information



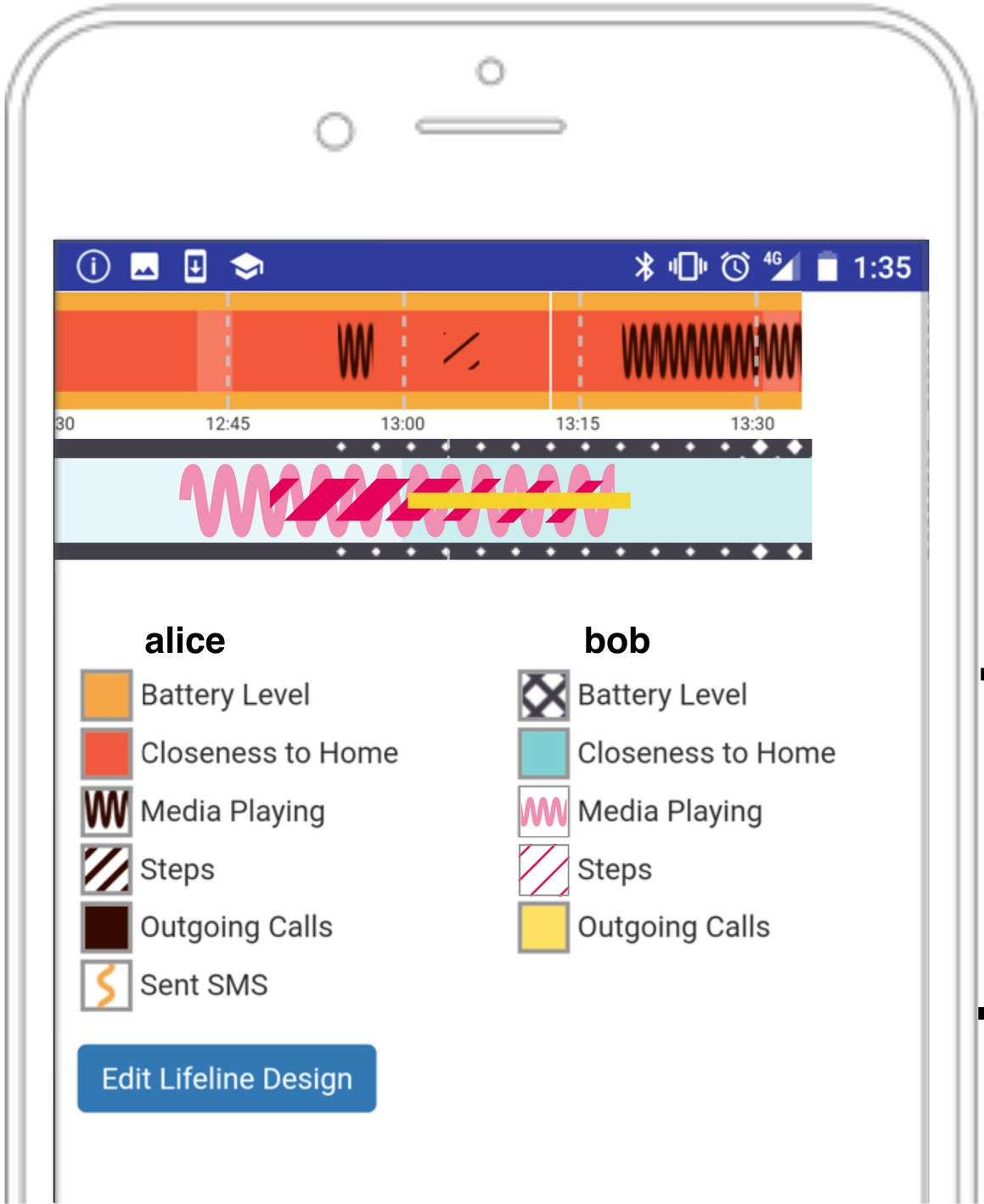
6 streams of contextual information



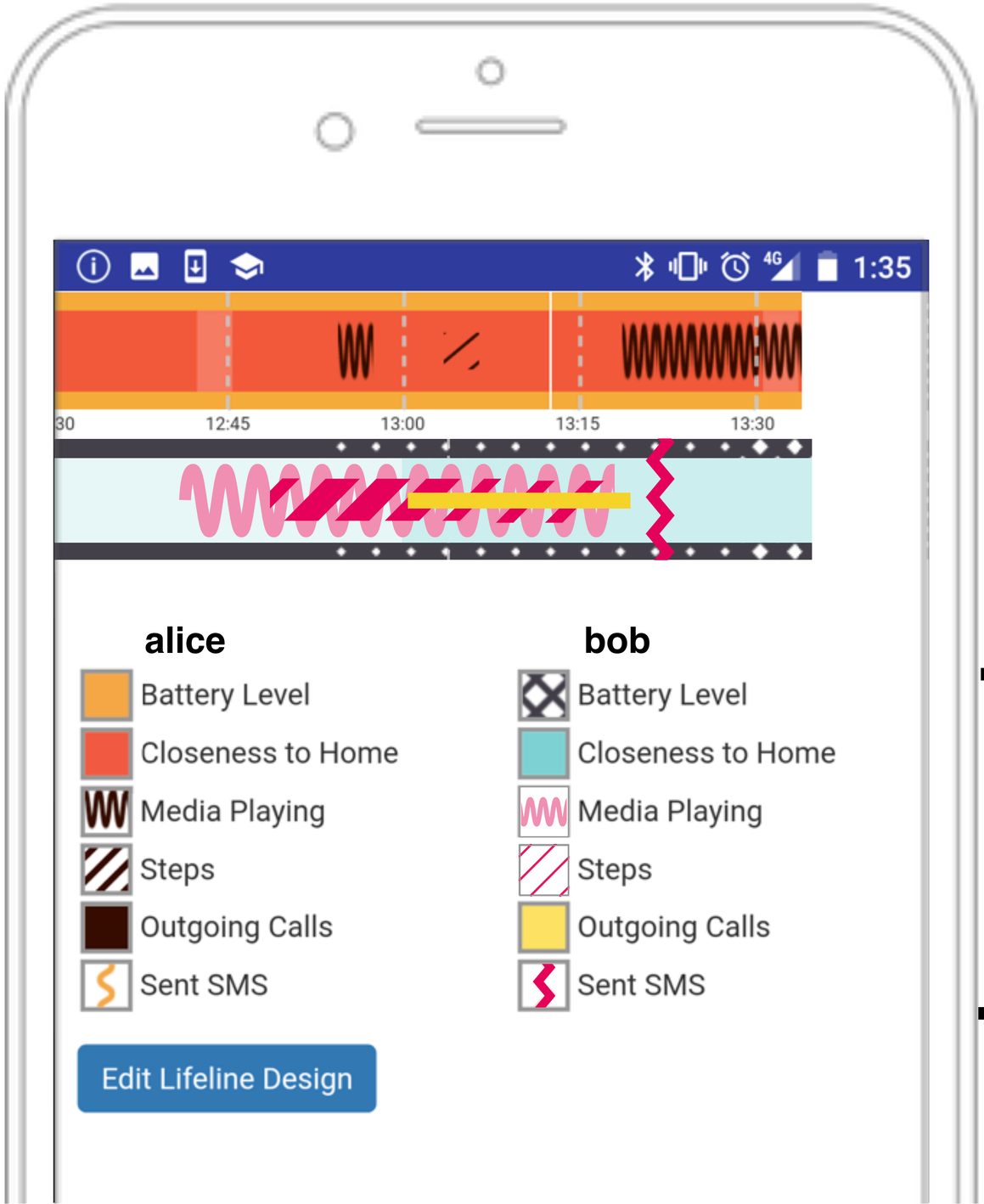
6 streams of contextual information



6 streams of contextual information



6 streams of contextual information



6 streams of contextual information


1 hour ago
now
18:45
19:00
19:15
19:30
His ***battery level*** stream
gets thinner as the phone's battery drops

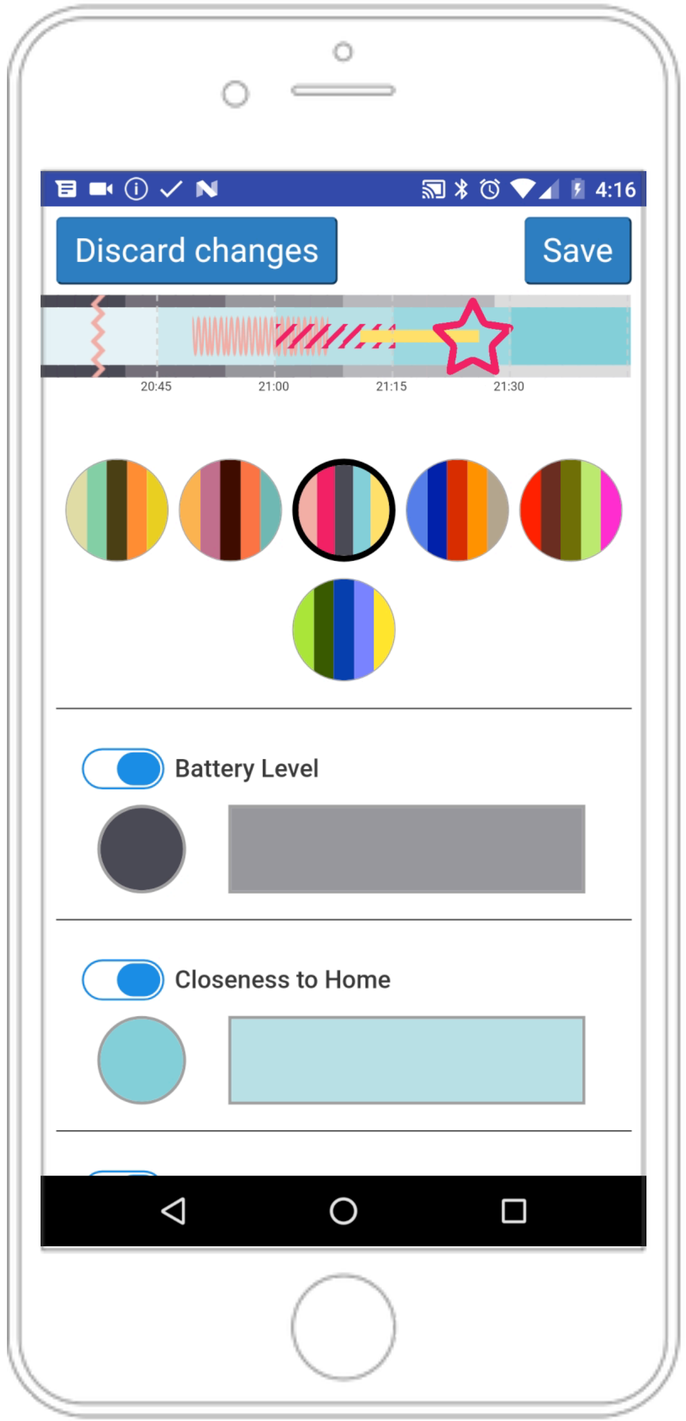
Far from home

Closer to home

The ***closeness to home*** stream
gets darker as the user gets closer to home

Very close
to home

The ***closeness to home*** stream
gets darker as the user gets closer to home



Partners individually choose
the streams they want to share with the other

Research goals

1) **Understanding adoption and real-world use of a new system:**
   - How do couples integrate Lifelines into their everyday communication?
   - How does this particular design helps partners understand each other's context?
     - Multiple streams of contextual information
     - Persistent streams of contextual information

2) **Discovering new opportunities for design:**
   1) What aspects of sharing contextual information can inspire new communication technologies for couples? Or for other relationships?
   2) Is there any aspect of this design that should be avoided in the future?

Considerations when designing field studies

1) How can we collect data of what is going on all the time, if we're not there to observe?
2) How can we get good data about user's experiences with our technology, if we're not there when the interesting events happen?
3) How do we counter novelty effects? For how long should we collect data?
4) How do we collect data without invading users' privacy?
5) What do we want to control, and what do we NOT want to control?

We ran a field study to answer these questions. Based on your own intuition, how would you design this field study?

METHOD

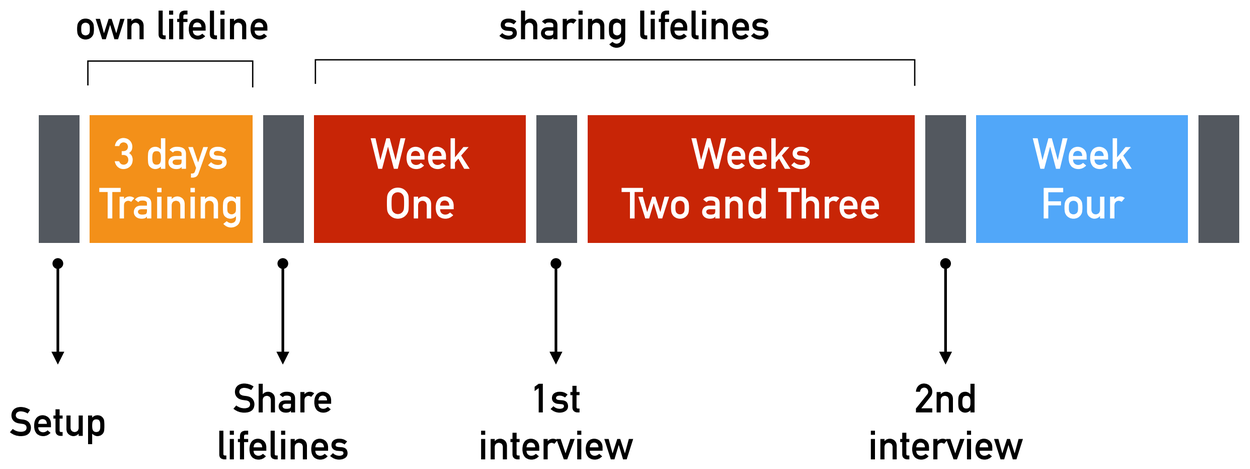
One-month technology-probe study

9 couples (8 living together)
Argentina, Brazil, UK, Switzerland, New Zealand & France

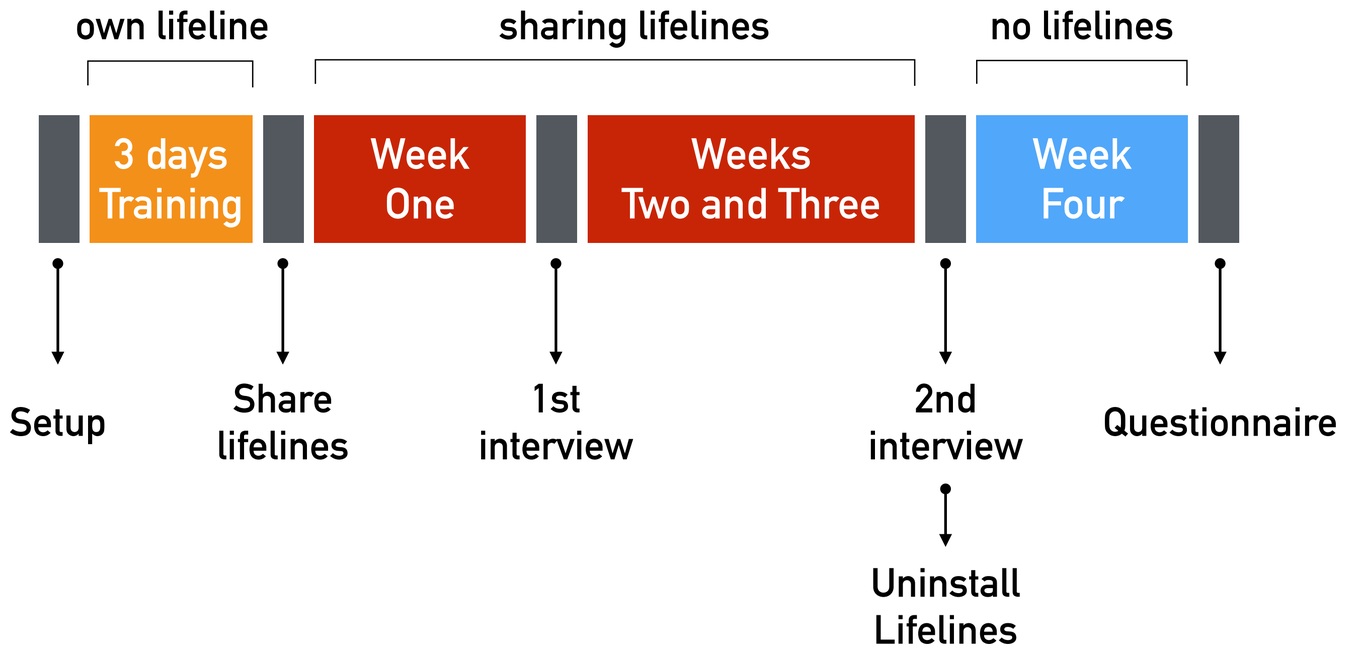
Setup and Interviews via Skype

During this “practice" period, participants can get familiar with how their context is visualised, and make informed decisions about what streams they want to share with their partners

We have a 1st interview after 1 week of sharing the contextual data so we can capture participant's first impressions, taking advantage of novelty effects



Then we wait 2 more weeks for the 2nd interview, to learn about their experiences past the novelty effects



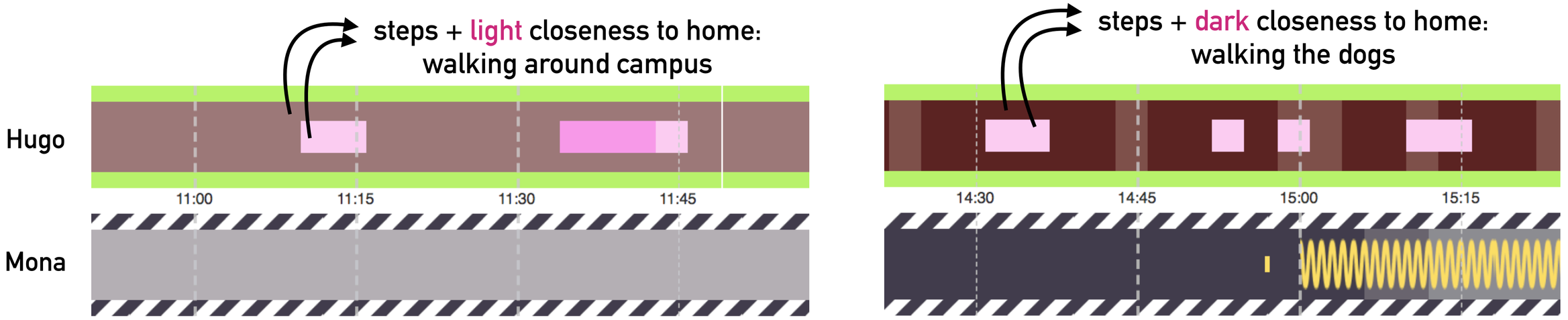
Last, we ask them to STOP sharing their lifelines and come back to them after a week, so we learn about the contrast of sharing vs. not sharing contextual information

We collected 22 hours of video from 34 Skype interviews

Open coding on videos (Chronoviz) + Thematic Analysis

During interviews, we show them their “old" data to help them remember interesting events and let them explain to us what happened in their own words

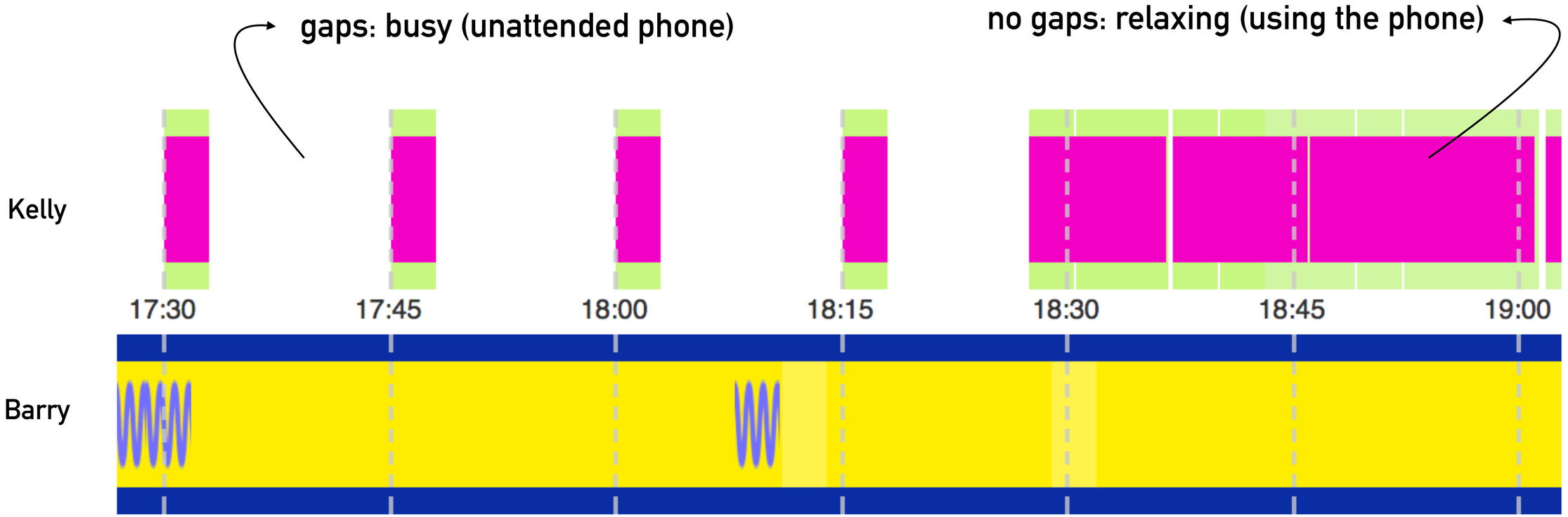
How did multiple streams inform context?



Interpreting multiple streams together helps disambiguate the meaning of single streams

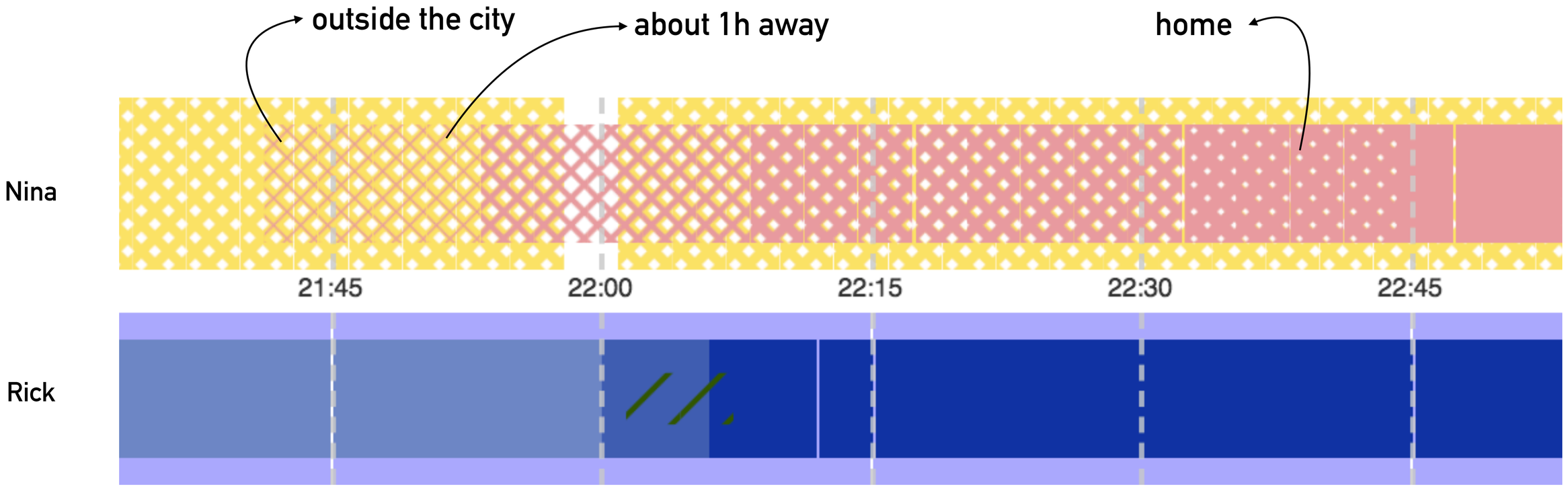
Participants most often looked at one stream at a time
according to the data that served their current needs

How did persistent streams inform context?



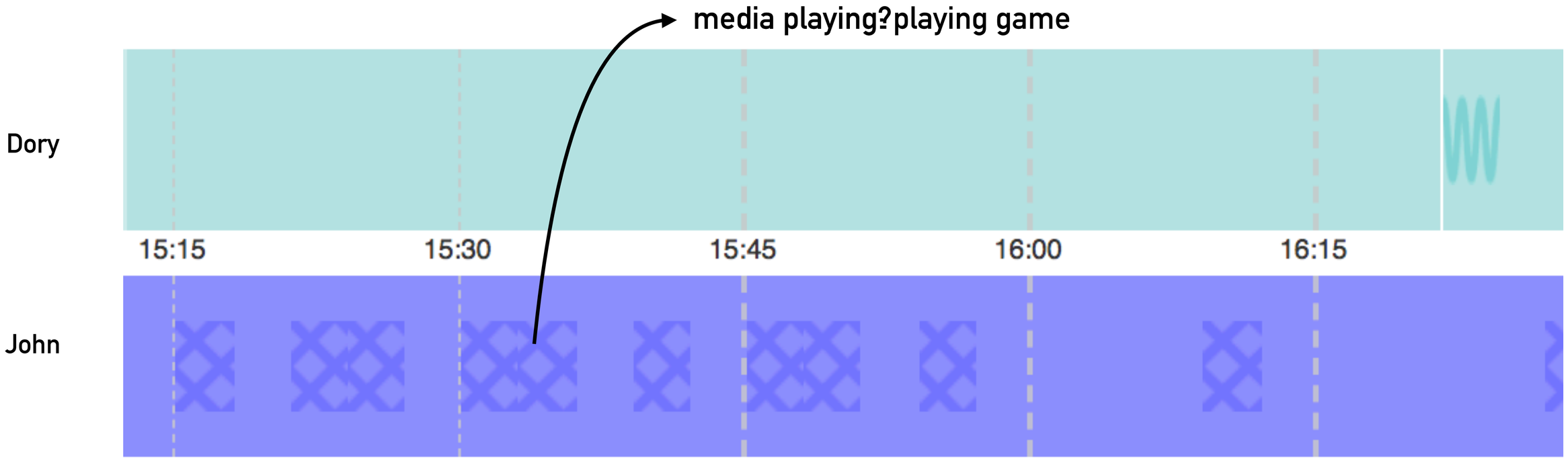
Patterns of missing data provide additional contextual information

How did couples' communication dynamics change?



Contextual information replaced direct communication

How did couples' communication dynamics change?



Contextual information triggered direct communication

RESULTS

Replacing direct communication
may hurt established communication patterns

Ambiguous visualizations enable plausible deniability

A poor balance between triggering and replacing
direct communication can lead to feeling more distant

Strong individual differences across and within couples

The streams that triggered or replaced communication depended on the routines, needs and intimate knowledge of each couple

	Media Playing	Steps
For Dory:	ask about surprising patterns (trigger)	check that her husband arrived to work safely (replace)
For Kelly:	coordinate pick-ups (replace)	warn her husband about his knee health (trigger)

Data that confirms partners' expectations and knowledge of each other replaces direct communication

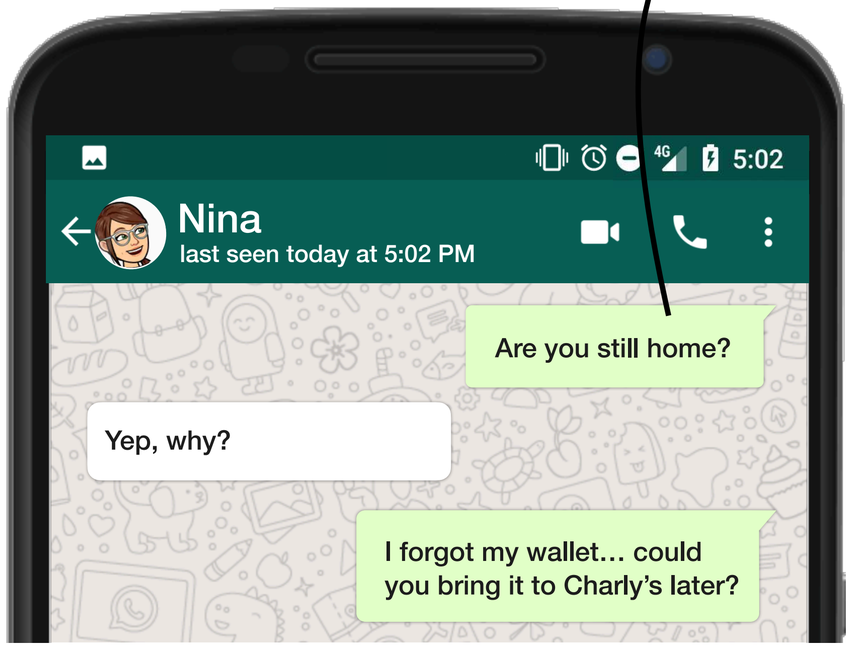
Data that challenges partners' expectations and knowledge of each other triggers direct communication

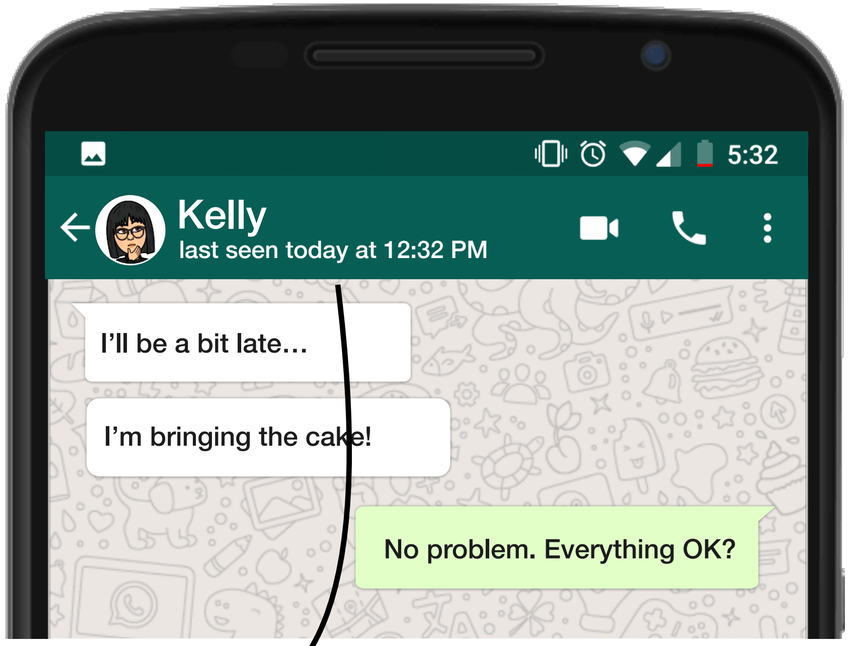
Post-Lifelines

15 out of 18 participants missed sharing at least 2 streams.

Some found ways of compensating the lack of Lifelines:

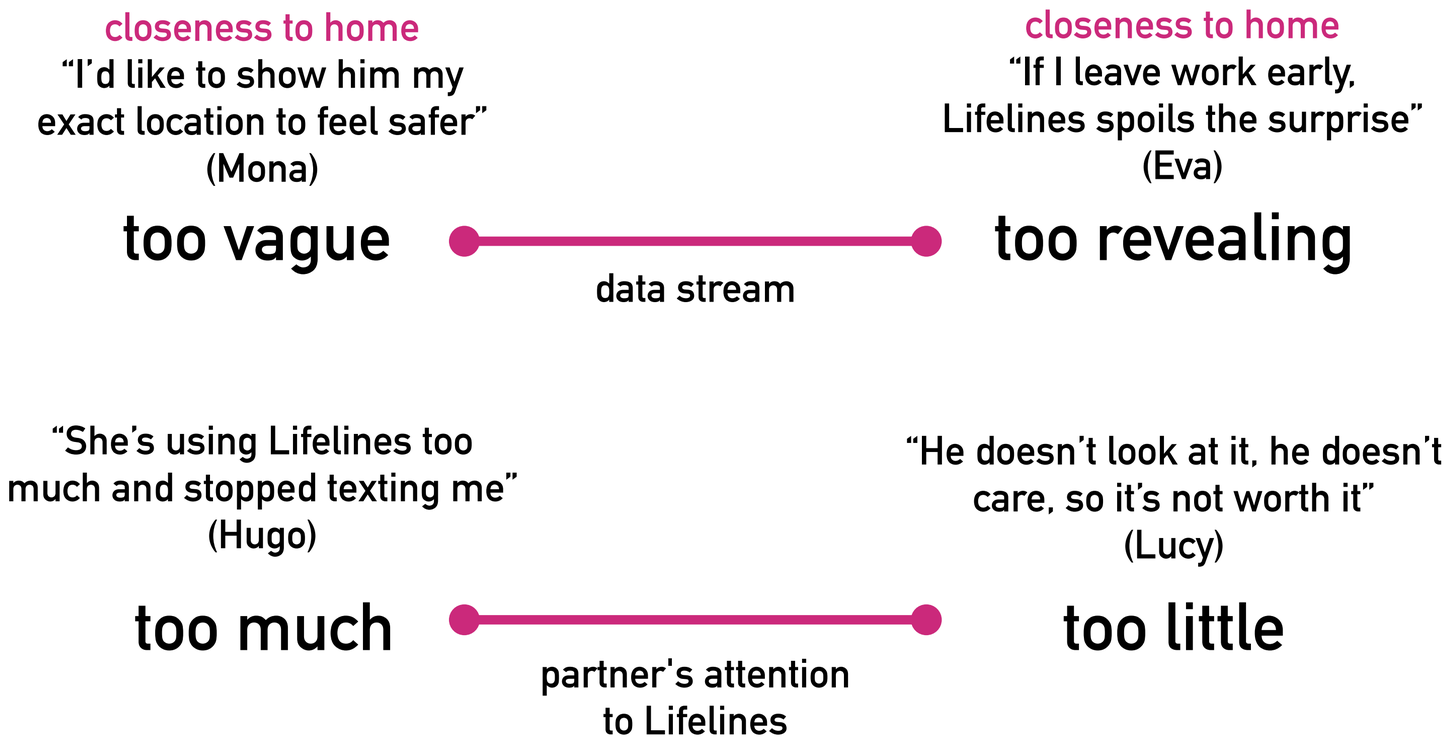
ask “where are you?” or “what are you doing?” before anything else (Rick)





check WhatsApp “last seen” status more often (Barry)

Individual differences around privacy concerns



Opportunities for design

Based on the strong individual differences in **privacy concerns** across couples: how about **customizable levels of persistence**? Some might prefer live, ephemeral visualisations rather than historical, persistent visualisations.

Based on the strong individual differences **in what data is meaningful** across couples: what if every app had a “context API” for enabling visualisations of more diverse types of contextual data?

Systems sharing contextual information should support **plausible deniability** to protect user’s privacy