




Inbox

University of Glasgow

- Horatiu Bota 
- ▶ **Self-Es: The Role of Emails-to-Self...** 13:30
- Finish slides - Think of good starting joke...

Microsoft Research

Paul N. Bennett @ MSR

- ▶ WHERE ARE YOUR SLIDES?! 13:29
- ...?

Ahmed H. Awadallah @ MSR

- ▶ Let us know how it goes! 11:12
- (No message text)

Susan T. Dumais @ MSR

- ▶ Good Luck... 10:33
- ...with the talk!

Self-Es: The Role of Emails-To-Self in Personal Information Management



Horatiu Bota

Today, 13:30

Horatiu Bota ▾

- Finish slides
- Think of good starting joke about closeness to lunch
- Email Paul, Ahmed and Susan about how it all went

University
of Glasgow

Microsoft®

Research



To/From same account
Single recipient
No **CCs**



Horatiu Bota

Today, 13:30

Horatiu Bota ▾

- Finish slides
- Think of good starting joke about closeness to lunch
- Email Paul, Ahmed and Susan about how it all went





User assistance:

- proactive display of Self-Es
- preferential ranking of Self-Es
- create UI elements that track Self-Es
- integration with existing tools (e.g. flags and folders)
- better support for task management (e.g. reminder Self-Es)



(RQ1) How **many** users send Self-Es?
How **often** do they send Self-Es?

(RQ2) **Why** do users email themselves?

(RQ3) Can **reminder** intent be **detected** in Self-Es?



(RQ1) How **many** users send Self-Es?
How **often** do they send Self-Es?

(RQ2) **Why** do users email themselves?

(RQ3) Can **reminder** intent be **detected** in Self-Es?



(RQ1) How **many** users send Self-Es?
How **often** do they send Self-Es?

(RQ2) **Why** do users email themselves?

(RQ3) Can **reminder** intent be **detected** in Self-Es?



(RQ1) How **many** users send Self-Es?
How **often** do they send Self-Es?

(RQ2) **Why** do users email themselves?

(RQ3) Can **reminder** intent be **detected** in Self-Es?



Avocado Email Collection

- Public email corpus ~2000s
- Sampled 88 active users
- Total of 110K sent emails

Pros: publicly available data

Cons: data from ~17 years ago



Self-E Survey

- ~20 questions, 4 section
- Distributed to MS employees (based in the US only)
- 238 respondents

Pros: reflects current behaviour

Cons: self-reported behaviour



Self-E Labelling Tool

- Website where users could label and donate their own Self-Es
- Accessible to MS employees
- 1274 donated Self-Es

Pros: user labelled Self-Es

Cons: privacy concerns





Avocado Email Collection

- Public email corpus ~2000s
- Sampled 88 active users
- Total of 110K sent emails

Pros: publicly available data

Cons: data from ~17 years ago



Self-E Survey

- ~20 questions, 4 section
- Distributed to MS employees (based in the US only)
- 238 respondents

Pros: reflects current behaviour

Cons: self-reported behaviour



Self-E Labelling Tool

- Website where users could label and donate their own Self-Es
- Accessible to MS employees
- 1274 donated Self-Es

Pros: user labelled Self-Es

Cons: privacy concerns





Avocado Email Collection

- Public email corpus ~2000s
- Sampled 88 active users
- Total of 110K sent emails

Pros: publicly available data

Cons: data from ~17 years ago



Self-E Survey

- ~20 questions, 4 section
- Distributed to MS employees (based in the US only)
- 238 respondents

Pros: reflects current behaviour

Cons: self-reported behaviour



Self-E Labelling Tool

- Website where users could label and donate their own Self-Es
- Accessible to MS employees
- 1274 donated Self-Es

Pros: user labelled Self-Es

Cons: privacy concerns





Avocado Email Collection

- Public email corpus ~2000s
- Sampled 88 active users
- Total of 110K sent emails

Pros: publicly available data

Cons: data from ~17 years ago



Self-E Survey

- ~20 questions, 4 section
- Distributed to MS employees (based in the US only)
- 238 respondents

Pros: reflects current behaviour

Cons: self-reported behaviour



Self-E Labelling Tool

- Website where users could label and donate their own Self-Es
- Accessible to MS employees
- 1274 donated Self-Es

Pros: user labelled Self-Es

Cons: privacy concerns



Email content

Self-E 1 / 53

Subject: after vacation
From: a.a.author@acm.org
To: a.a.author@acm.org
Sent: 09/02/2016 at 2:40PM

No attachments

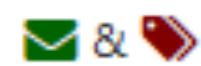
- + Set meeting with Sara
- + Check whether John resolved the dependency
- + Ask Pat to pushback timeline



Skip
this email



Submit email
labels only



Submit email
content and **labels**

Email labels

This email is a:

(check all that apply)

[Expand all](#) / [Close all](#)

▼ Reminder or ToDo

- Thing(s) to do
- Thing(s) to buy
- Thing(s) to remember
- Other

▼ Transfer

- Across devices
- Across accounts
- Archive or backup
- Other

▼ Other

- Test email (e.g. check email works)
- Draft of a message to send later
- Forward to myself of other email
- Sent this to myself by mistake
- This is NOT a Self-E
- Other

▼ Copy/Paste

- Link(s) or URL(s)
- Photo(s) / Image(s)
- Quote(s)
- Travel directions
- Other directions (e.g. recipes)
- Phone number(s)
- Other

▼ Event note(s)

- Before meeting or event
- During meeting or event
- After meeting or event
- Other

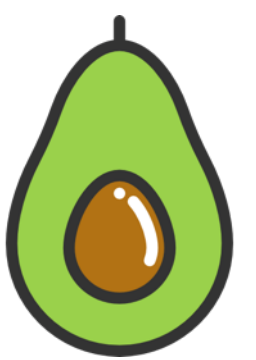
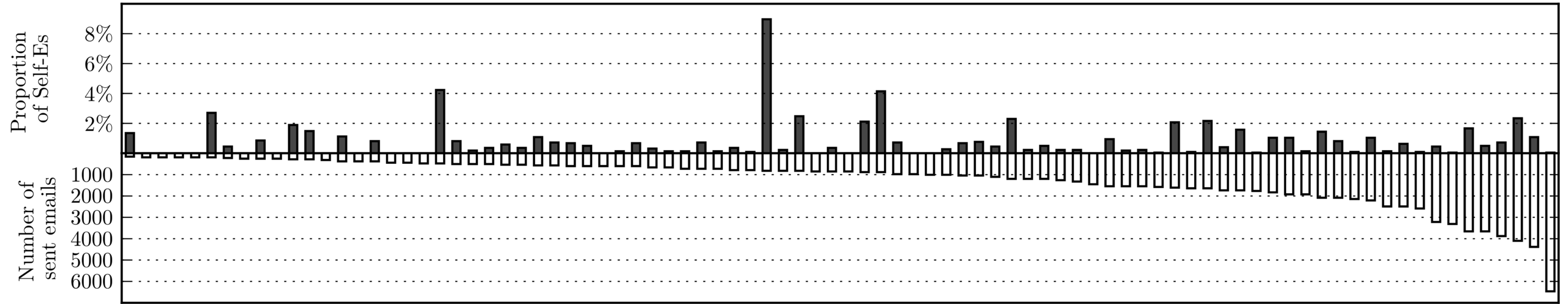
This e-mail is / was useful for a few:

N/A Never Hours Days Weeks Months Years

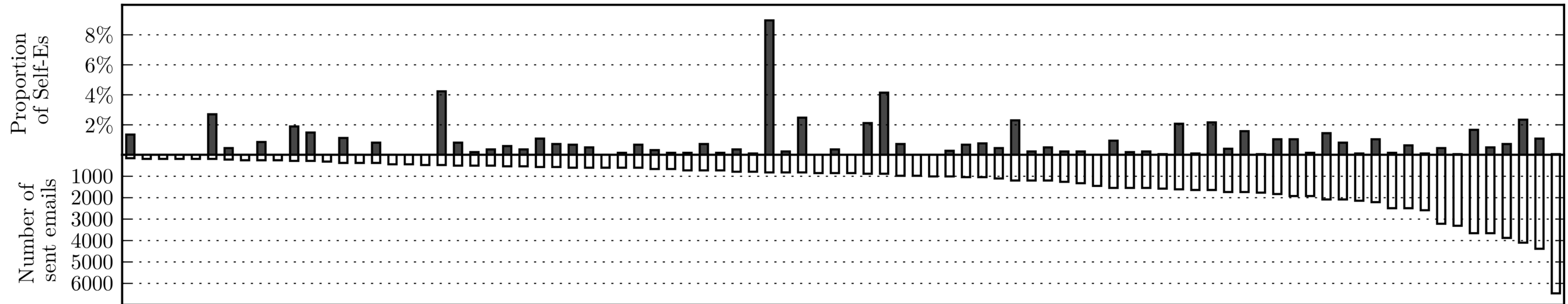


Self-E labelling tool

Results: (1) How many users send Self-Es?

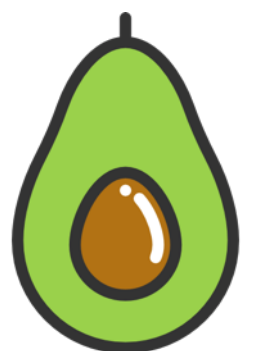


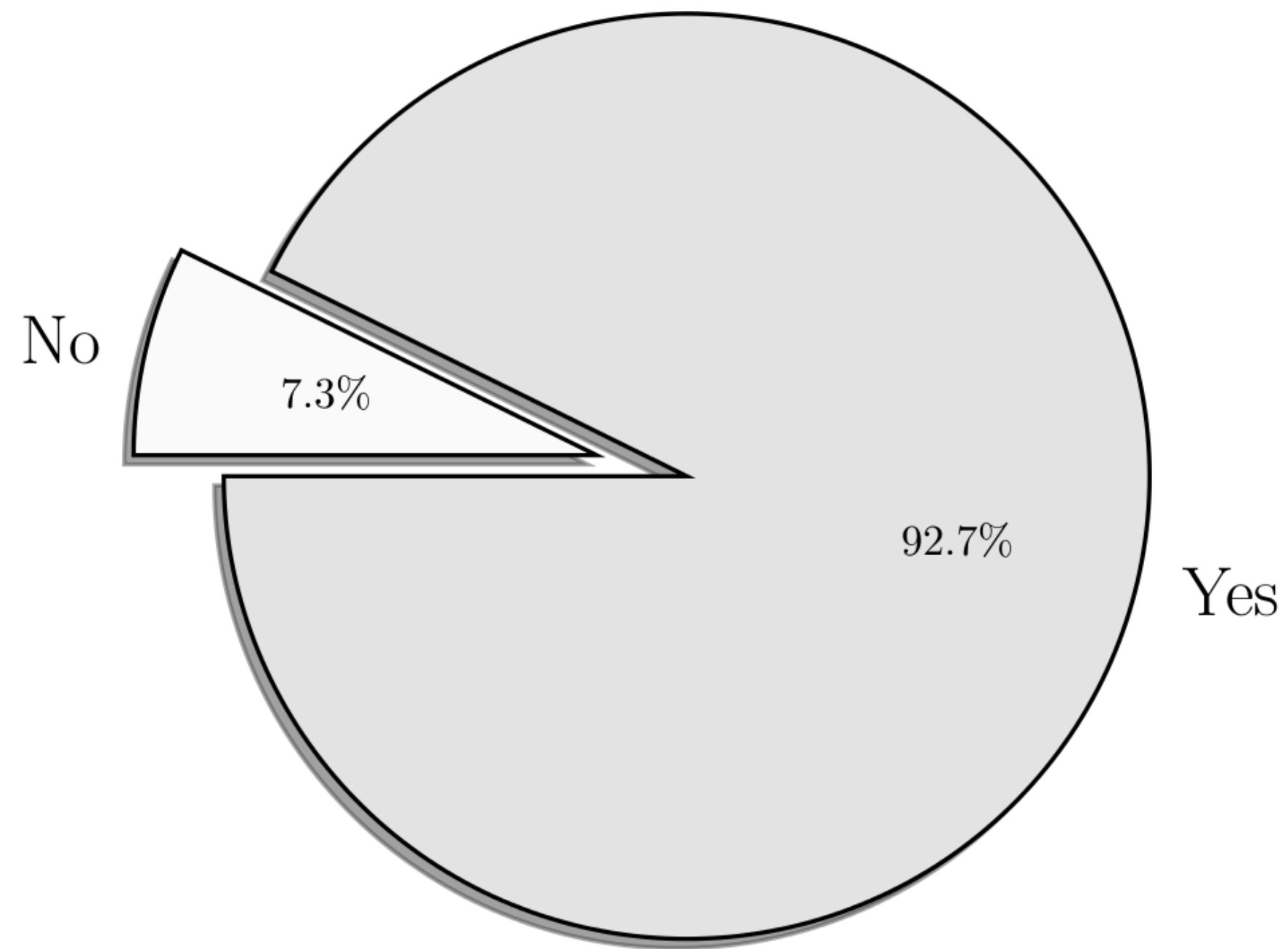
Results: (1) How many users send Self-Es?



In the Avocado email corpus:

- >80% users have sent at least one Self-E
- For ~75% users, <1% of sent emails are self-addressed
- Highest proportion: ~8% **Self-Es** in ~1000 sent emails





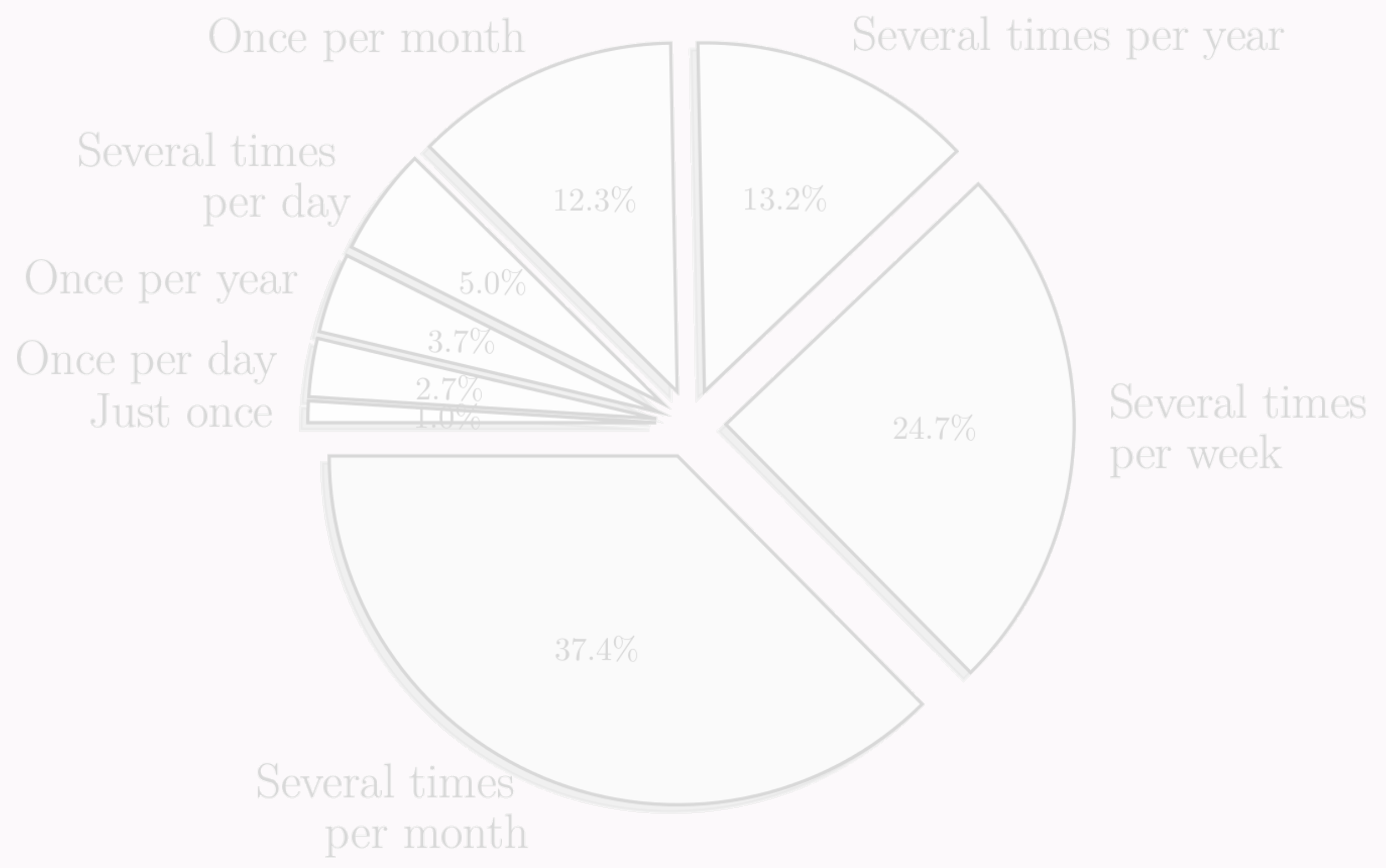
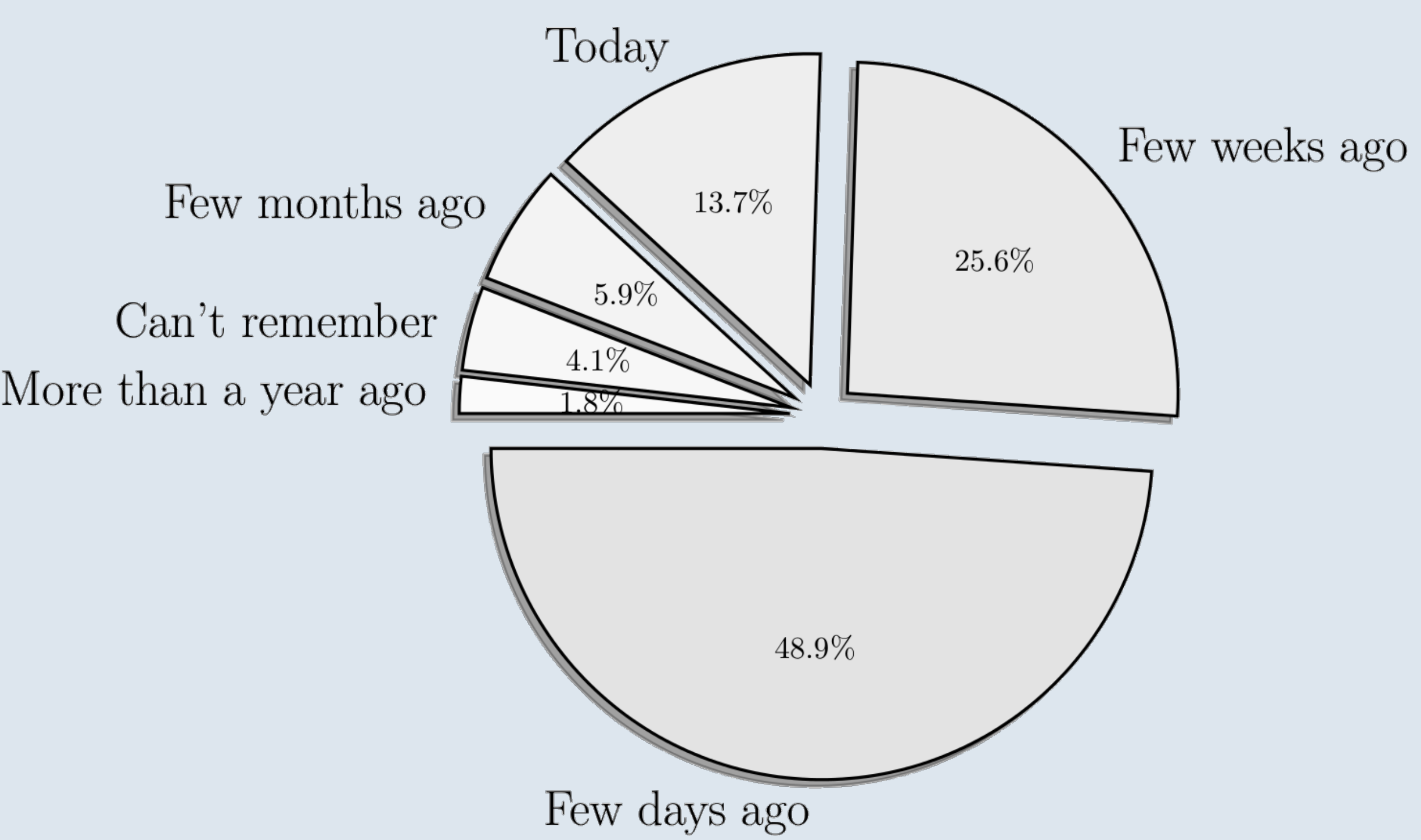
Overall:

- 80% - 90% of email users have sent at least one Self-E

(a) Survey: Have you ever sent an email to yourself, and only yourself, containing information that is useful to you?



Results: (1) How often do they send Self-Es?

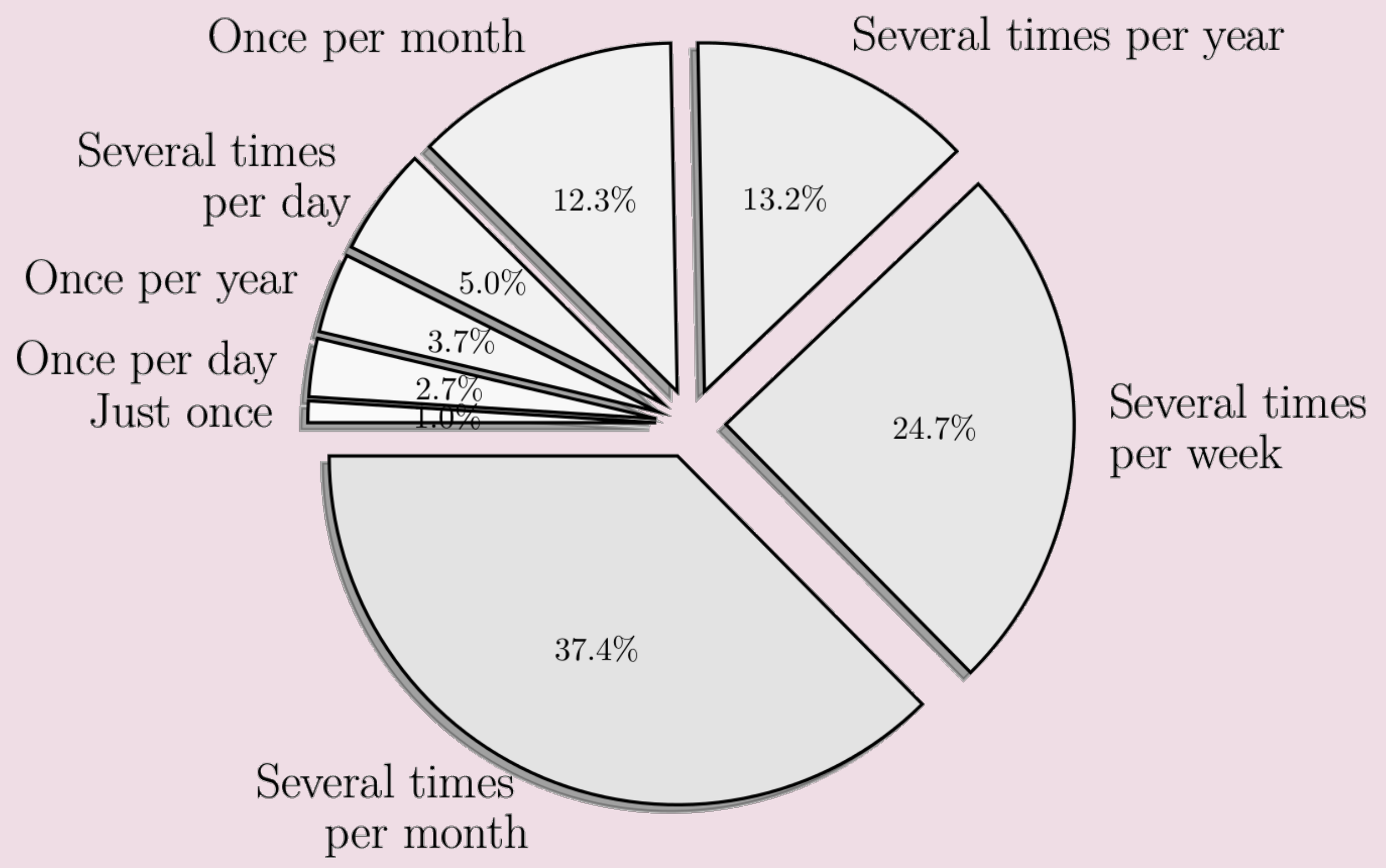
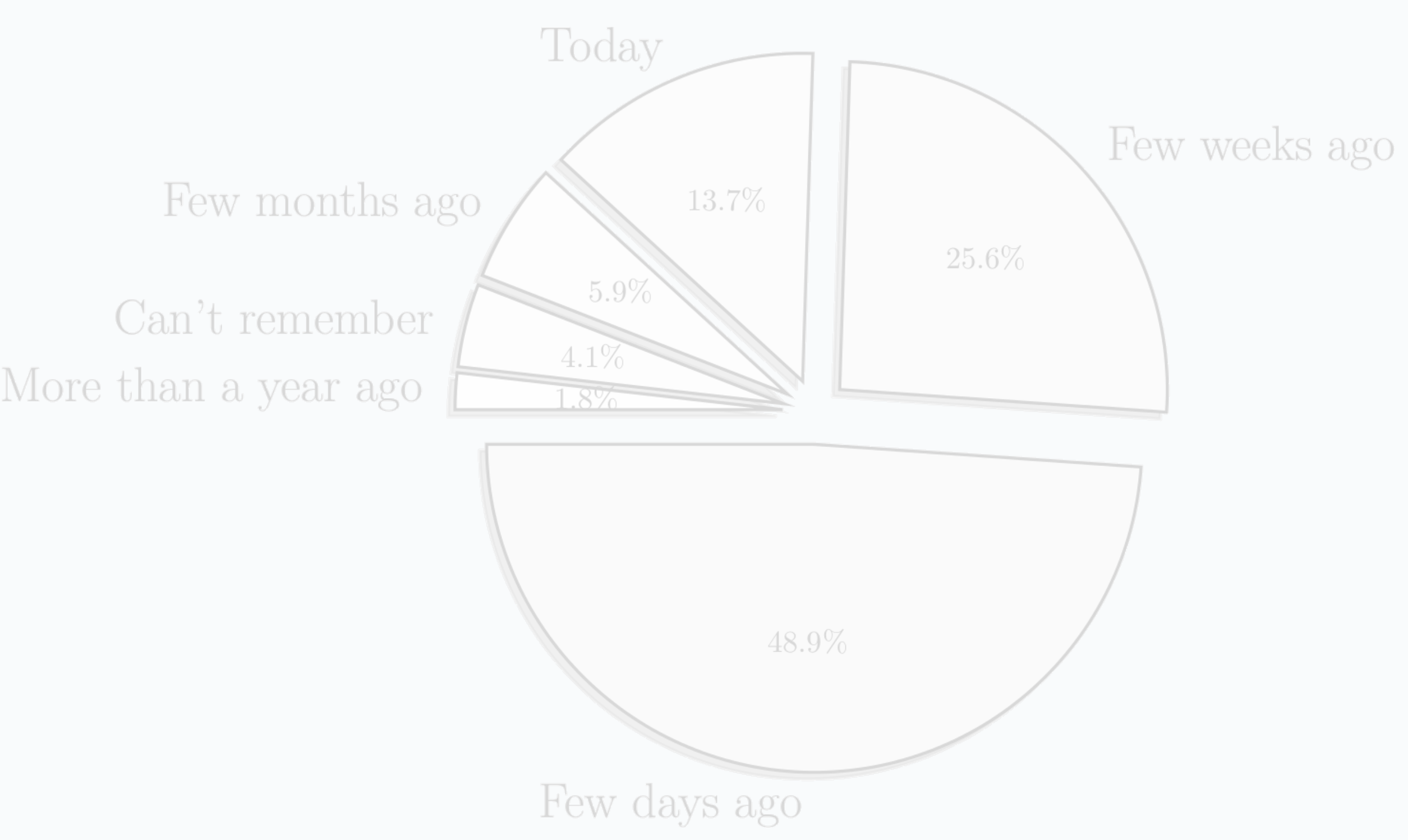


Most recent Self-E:

- ~75% sent a Self-E a "Few weeks ago" or "Few days ago"



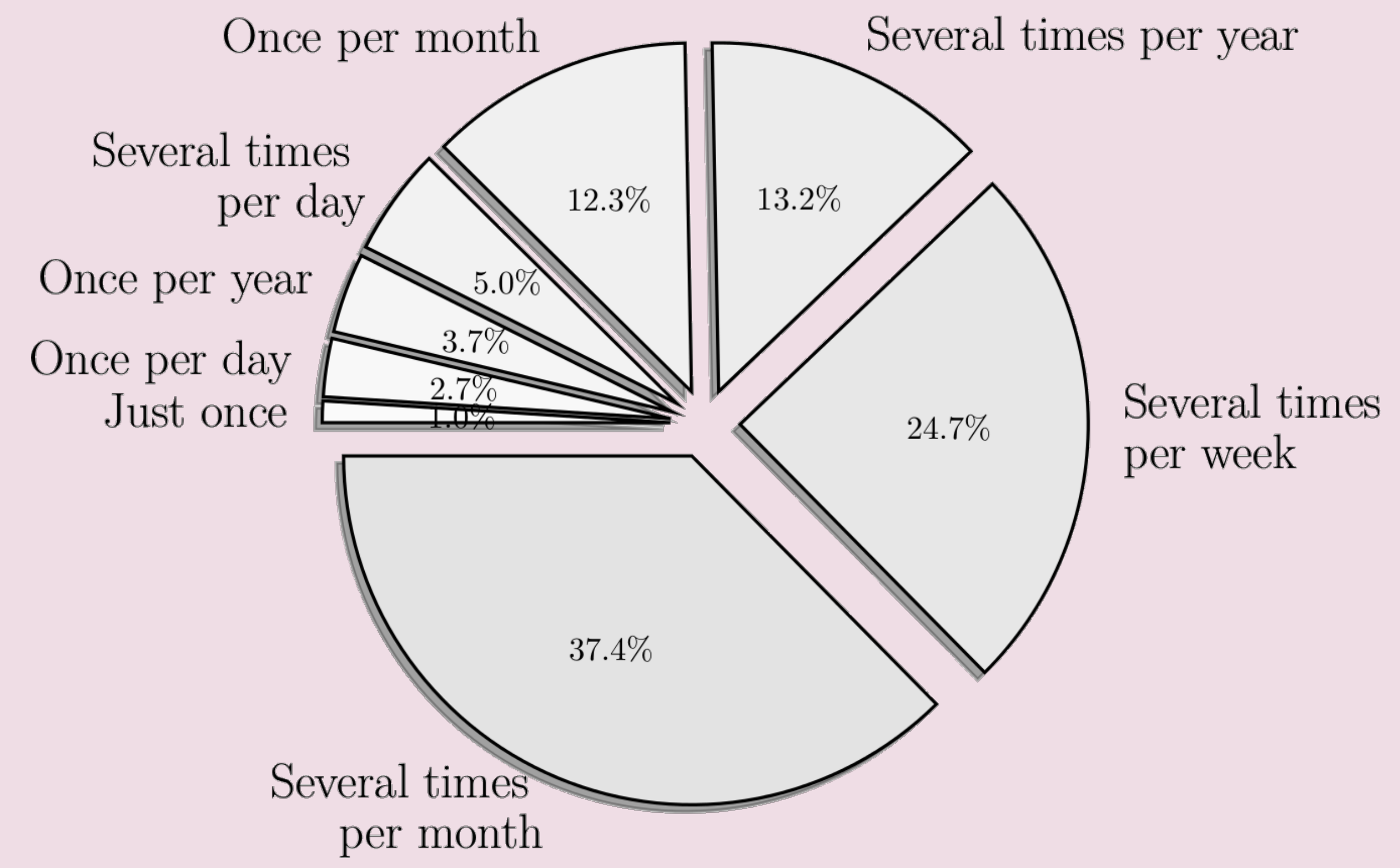
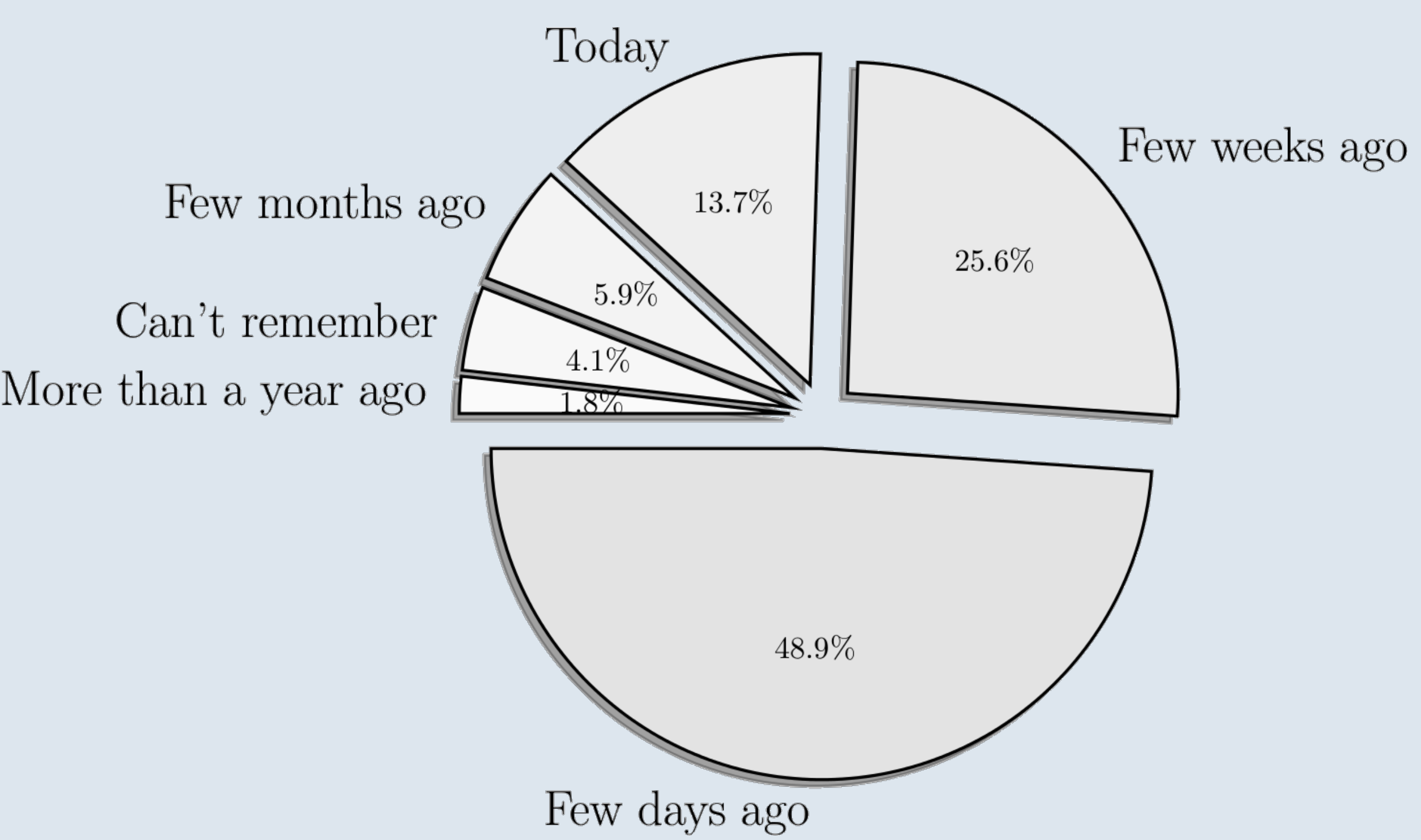
Results: (1) How often do they send Self-Es?



How often?

- ~75% send Self-Es "Several times per month" or more frequently

Results: (1) How often do they send Self-Es?



Most recent Self-E:

- ~75% sent a Self-E a "Few weeks ago" or "Few days ago"

How often?

- ~75% send Self-Es "Several times per month" or more frequently

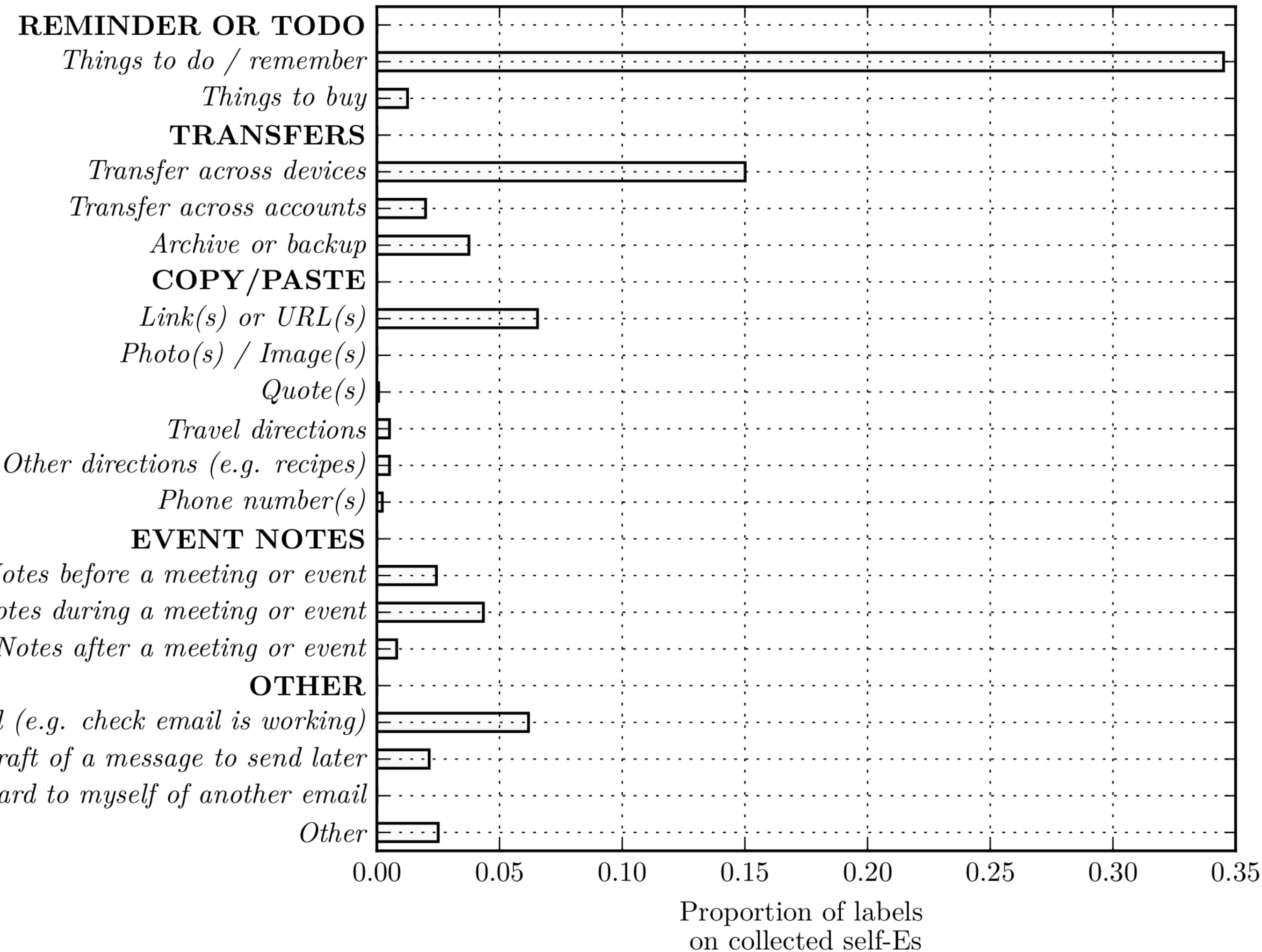
How many users
send Self-Es?

**80% - 90% of email users
have sent at least one Self-E**

How often do they
send Self-Es?

**3/4 send Self-Es several
times per month**

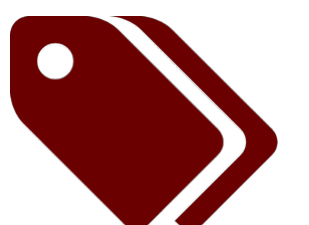
**1/3 send two or more
Self-Es per week**



Labelled Self-Es:

- **Reminder or TODO, File transfers, Links or URLs** are the most popular labels
- **53%** of donated Self-Es were **Reminder or TODOs**

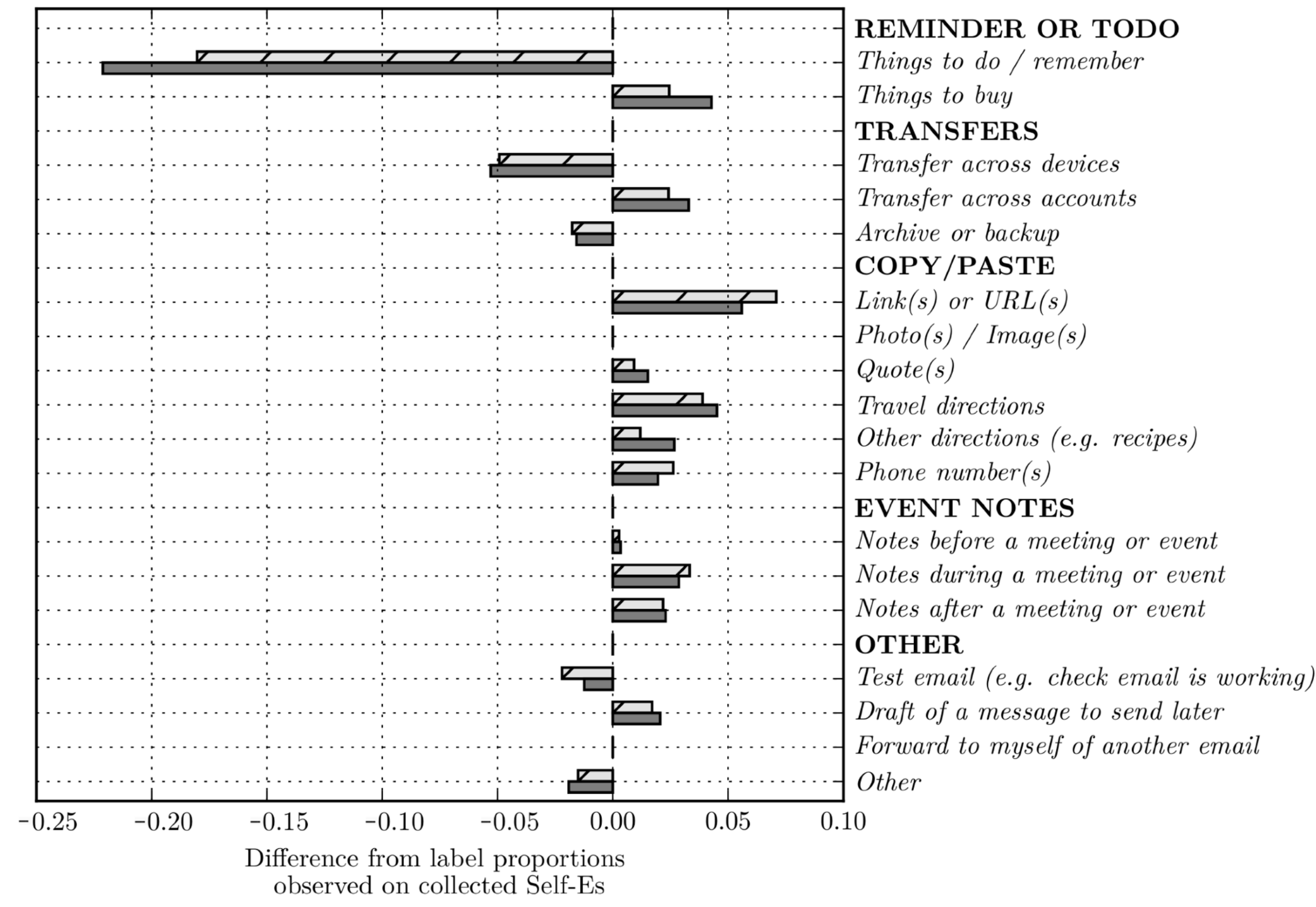
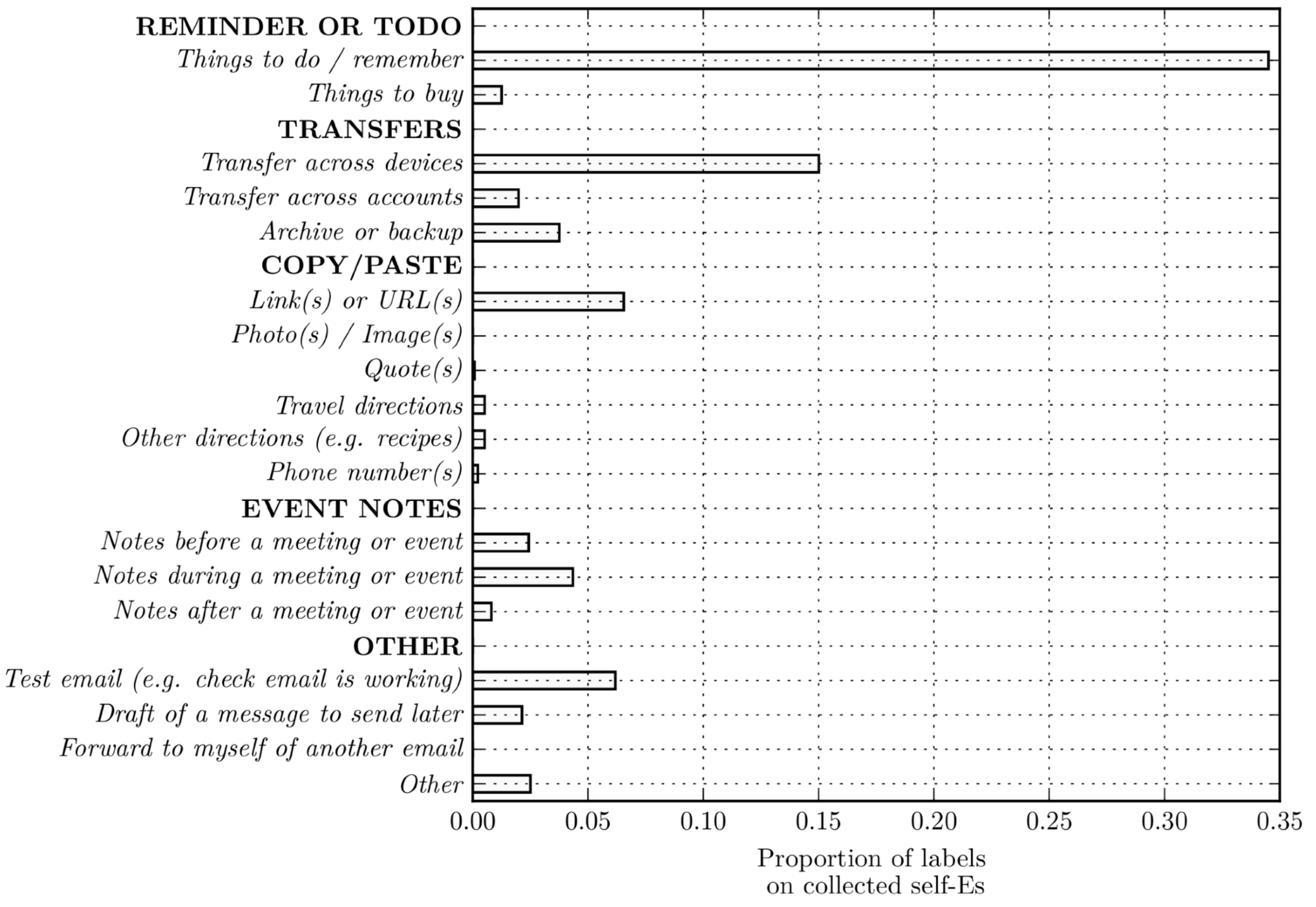
- **Other** includes: test emails, passwords, code snippets or photos of receipts



Results: (2) Why do users email themselves?

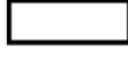




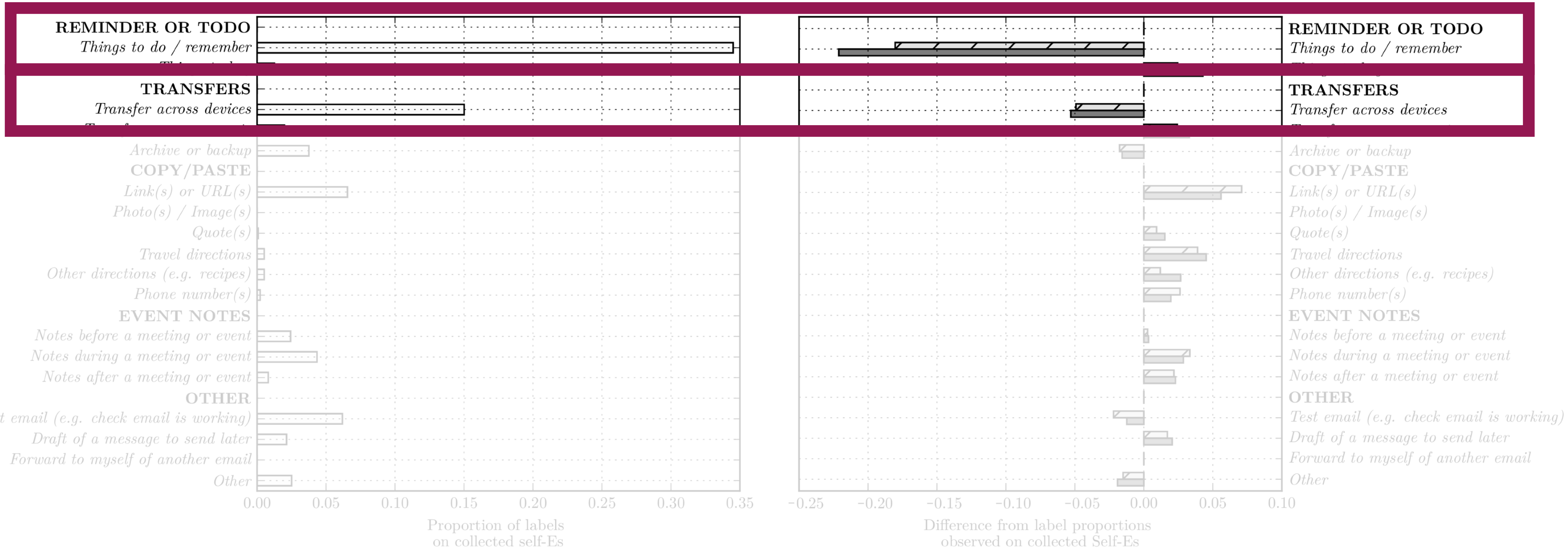
- Tool: Labels collected through the Self-E labelling tool
- Survey: What type of information was contained in your most recent Self-E?
- Survey: What types of information do you generally have in your Self-Es?



Results: (2) Why do users email themselves?



-  Tool: Labels collected through the Self-E labelling tool
-  Survey: What type of information was contained in your most recent Self-E?
-  Survey: What types of information do you generally have in your Self-Es?





Broad spectrum of intents

**50% of Self-Es have some
reminder / todo intent.**

| Feature set | Mean accuracy | |
|------------------------|-------------------|-------------------|
| | Flat | Stacked |
| All features | 0.651 (±0.046) | 0.786 (±0.041) |
| Content features only | 0.606 (±0.067) | 0.747 (±0.048) |
| Metadata features only | 0.619 (±0.074) | 0.641 (±0.057) |

"Flat" classifier

- simple logistic regression

"Stacked" classifier

- **predict multi-class labels:** Transfer, Copy/Paste, Notes, Other
- use **multi-class predictions as features** in the classification

| Feature set | Mean accuracy | |
|------------------------|-------------------|-------------------|
| | Flat | Stacked |
| All features | 0.651 (±0.046) | 0.786 (±0.041) |
| Content features only | 0.606 (±0.067) | 0.747 (±0.048) |
| Metadata features only | 0.619 (±0.074) | 0.641 (±0.057) |

"Flat" classifier

- simple logistic regression

"Stacked" classifier

- **predict multi-class labels:** Transfer, Copy/Paste, Notes, Other
- use **multi-class predictions as features** in the classification

| Feature set | Mean accuracy | |
|------------------------|--------------------------|--------------------------|
| | Flat | Stacked |
| All features | 0.651 (± 0.046) | 0.786 (± 0.041) |
| Content features only | 0.606 (± 0.067) | 0.747 (± 0.048) |
| Metadata features only | 0.619 (± 0.074) | 0.641 (± 0.057) |

"Flat" classifier

- simple logistic regression

"Stacked" classifier

- **predict multi-class labels:**
Transfer, Copy/Paste, Notes, Other
- use **multi-class predictions**
as features in the classification



How many users send Self-Es?

- **80% - 90%** email users sent at least one Self-E
- **75%** send several Self-Es / month
- **65%** send one or more Self-E / week

Why do users email themselves?

- **>50%** Self-Es are **reminders** or **todos**
- **File transfers across devices, Links or URLs, Test emails** are popular types of Self-Es

Can we detect reminder Self-Es?

- **78%** accuracy predicting Reminder intent
- **Text features + Email metadata** features



How many users send Self-Es?

- **80% - 90%** email users sent at least one Self-E
- **75%** send several Self-Es / month
- **65%** send one or more Self-E / week

Why do users email themselves?

- **>50%** Self-Es are **reminders** or **todos**
- **File transfers across devices, Links or URLs, Test emails** are popular types of Self-Es

Can we detect reminder Self-Es?

- **78%** accuracy predicting Reminder intent

- **Text features + Email metadata** features



How many users send Self-Es?

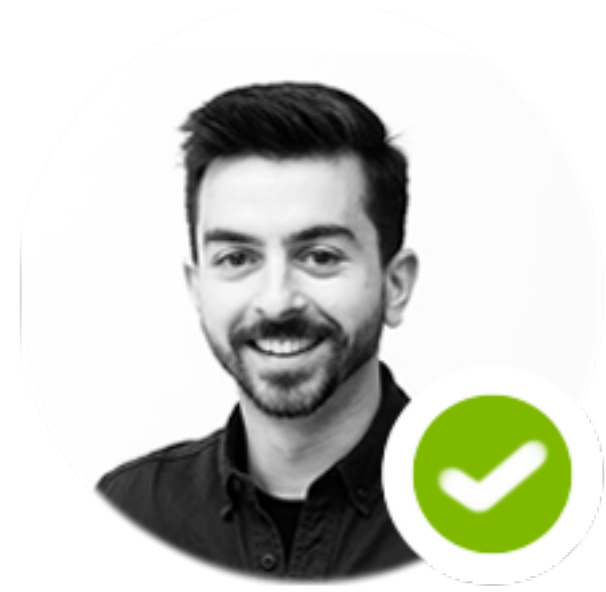
- **80% - 90%** email users sent at least one Self-E
- **75%** send several Self-Es / month
- **65%** send one or more Self-E / week

Why do users email themselves?

- **>50%** Self-Es are **reminders** or **todos**
- **File transfers across devices, Links or URLs, Test emails** are popular types of Self-Es

Can we detect reminder Self-Es?

- **78%** accuracy predicting Reminder intent
- **Text** features + **Email metadata** features



Horatiu

Ph.D. Student at University of Glasgow
www.horatiubota.com



Paul



Ahmed



Susan