**Appendices to:** Van Roekel, H., Giurge, L., Schott, C., & Tummers, L.Nudges can be both autonomy-preserving and effective. Evidence from a survey and quasi-field experiment. Behavioural Public Policy.

**Appendix A: Interview guide**

Context: Qualitative pre-study for quasi-field experiment.

Goal pre-study: Developing appropriate nudges to reduce work pressure and stress.

Specific goals: 1) Defining the behavioral problem. 2) Deciding whether choice architecture is appropriate. 3) Checking whether there are bottlenecks that may hamper intervention power.

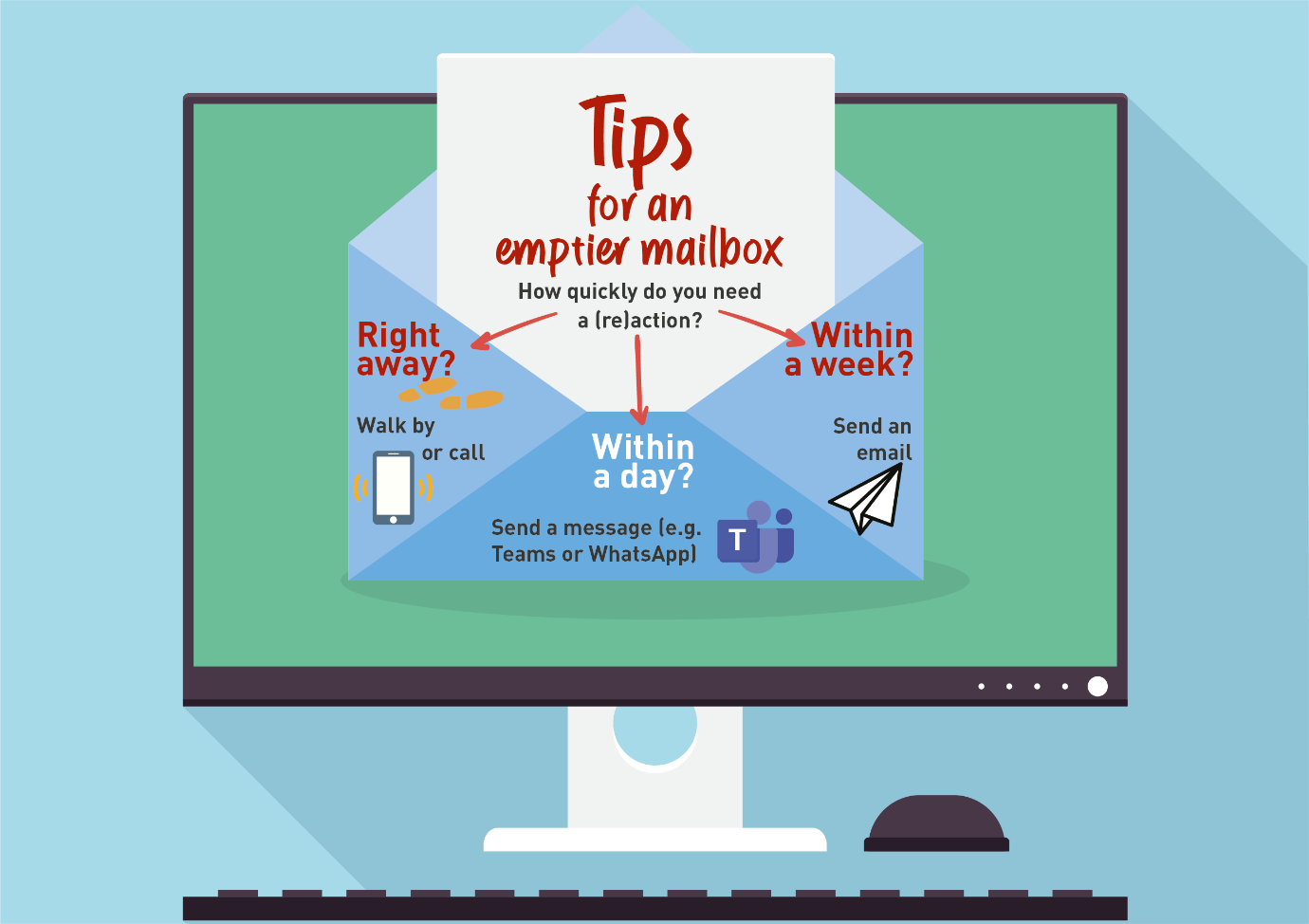
1. Before the interview
   1. Written informed consent to interview participation
2. Introduction
   1. Consent to record interview.
   2. Introducing the researcher.
   3. Explaining the research.
   4. Introducing the interviewee
      1. Who are you and what is your job title?
      2. How long have you been doing this?
3. Behavioral problem
   1. Do you have to deal with work pressure/stress in your work?
      1. In your own work?
      2. In work of colleagues/other employees?
         1. What groups of colleagues?
   2. In what concrete moments do you experience work pressure/stress? E.g. [partly inductively/deductively generated list]
      1. *Working overtime*
      2. *Working extra shifts*
      3. *Presenteeism*
      4. *Not taking holidays*
      5. *Taking no breaks*
      6. *Checking your work phone at home*
      7. *Not being able to say no to requests (to which requests?)*
      8. *Using email*
      9. *Meetings*
   3. What are the consequences hereof for you?
      1. For your work with clients.
      2. For your work with colleagues.
      3. For your private life.
   4. In what concrete situations do you observe work pressure/stress among your colleagues and what are the consequences hereof? For which groups of colleagues?
4. Goal behavior
   1. How would you or your colleagues behave ideally regarding the situations that you mentioned?
5. Why is there a problem
   1. What causes the gap between the goal behavior and the behavioral problem?
      1. Causes
      2. Causers
   2. Do you have ideas to reduce work pressure/stress regarding this situation?
6. Closing

**Appendix B: Nudges.**

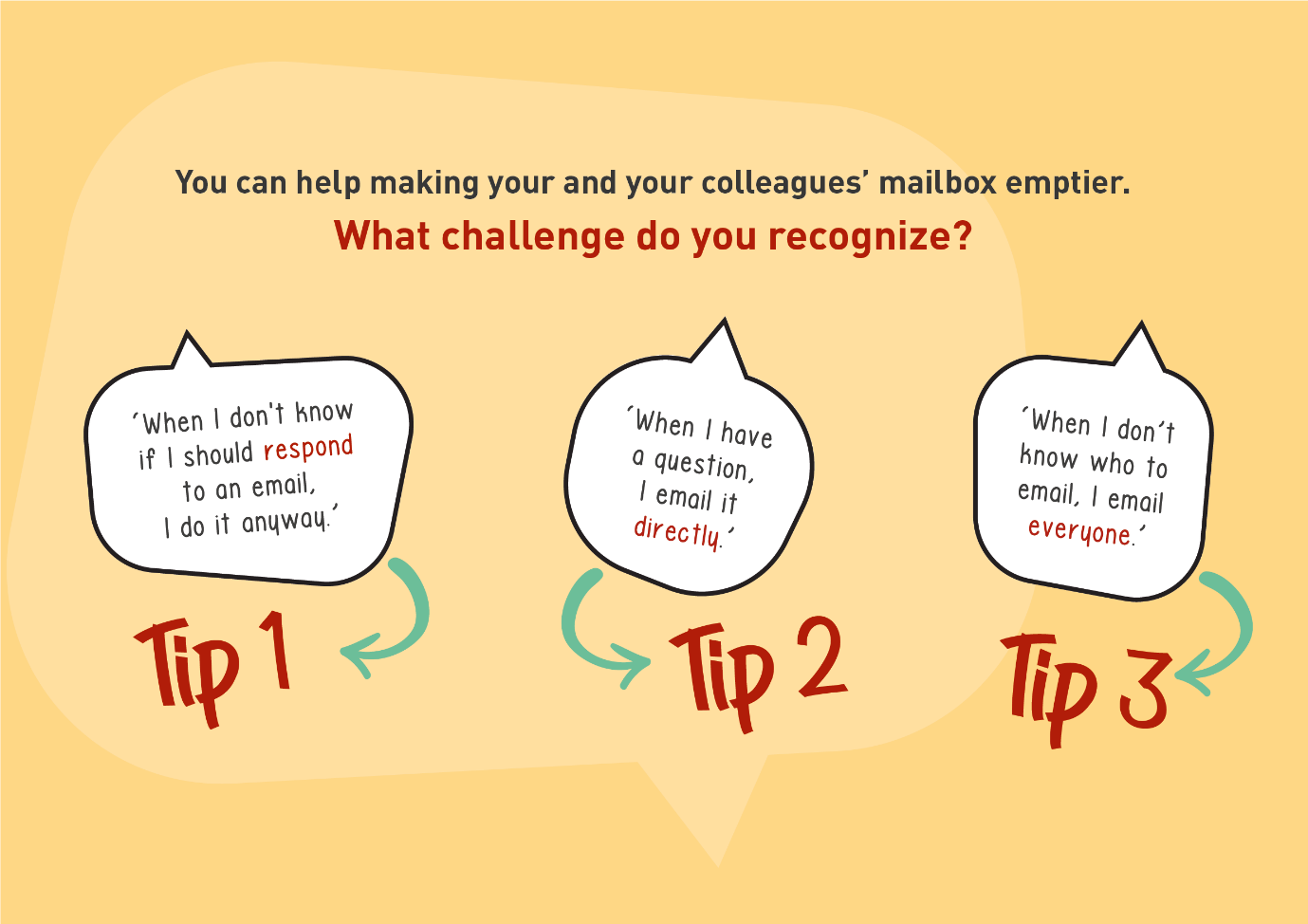
The first nudge, an opinion leader nudge, is a classic nudge that nevertheless is expected to preserve autonomy because it targets System 2 thinking by providing information. An opinion leader nudge is a message in which the behavior of a person of influence is described. Its expected effect is that receivers of this message will adapt this behavior due to the position of the opinion leader. The mechanism is based on two assumptions. First, in making decisions people rely on social reference points, i.e., the behavior of others (Münscher et al., 2016). For example, descriptive social norms, statements about what other people would do in a situation, stimulate people to conform. Furthermore, opinion leaders are highly respected messengers whose views or behaviors people are more likely to adapt. There are multiple ways of identifying such opinion leaders, but a common way to do so is by selecting those in formal leadership positions or with a specific expertise (Valente and Pumpuang, 2007). The opinion leader nudge in this study presented a message from ‘your HR manager’ saying that they notice emailing too much causes unnecessary stress and they are therefore going to email less, suggesting this will give more calmness in work, and asking to join them.

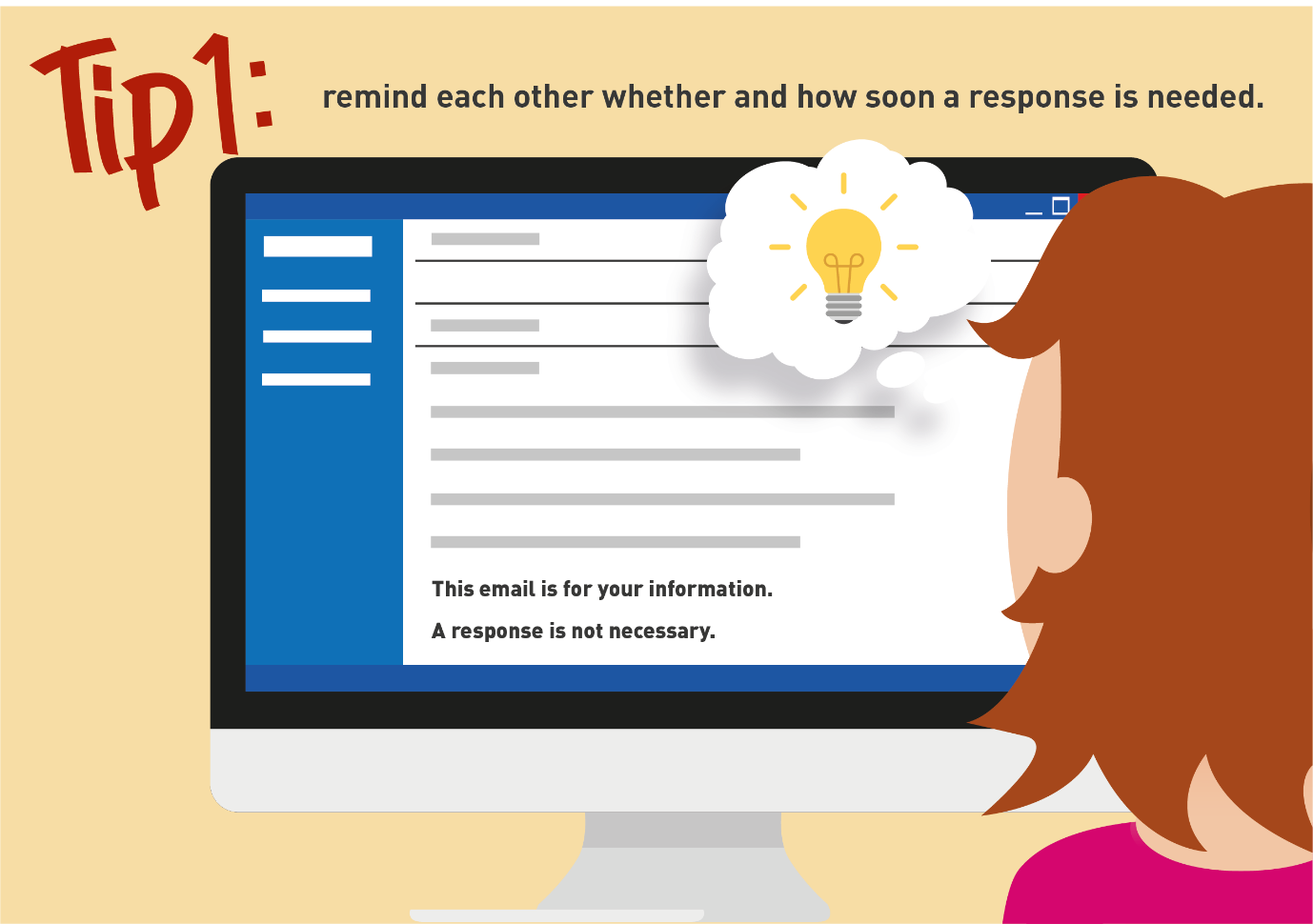
*Image 1: Opinion leader nudge.*

The second nudge, a rule-of-thumb, attempts to implement more conscious behavior change, resembling a boost. A rule-of-thumb nudge presents information in an understandable way so that less effort is required to make a decision (Hertwig and Grüne-Yanoff, 2017; Münscher et al., 2016). It is a hybrid behavioral intervention. Münscher et al. (2016) refer to a rule-of-thumb as a simplification nudge, but it also resembles a simple decision tree, which is a type of boost (Hertwig and Grüne-Yanoff, 2017, p. 979). The rule-of-thumb in this study provides a simple question to decide whether email is appropriate: how quickly do you need a (re)action? The suggested rule-of-thumb is that if you need it right away, you should walk by or call; if you need it within a day, you should send a message; and if you need it within a week, you should send an email.

*Image 2: Rule-of-thumb.*

Finally, we describe self-nudges, nudges to be used by employees themselves. The concept of self-nudging was recently introduced by Reijula and Hertwig (2022). Many nudges could be turned into self-nudges, the difference being that the person who is nudged is also the one who nudges. Enabling people to apply self-nudges can be seen as a type of self-control boost (Hertwig and Grüne-Yanoff, 2017, p. 979). Examples of self-nudges reminding yourself of a certain decision (e.g., by putting up a note on their computer screen) or adapting a different frame for the same decision (e.g., by thinking about working out at the gym as a privilege rather than a chore) (Reijula and Hertwig, 2022). In this study, three self-nudges were proposed. To create awareness among employees that they can influence their own choice architecture to change their behavior (Reijula and Hertwig, 2022), we introduced the self-nudges by saying ‘You can help making your and your colleagues’ mailbox emptier. What challenge do you recognize?’ After this, three challenges were introduced, each connected to one of the self-nudges. The challenges are: email response uncertainty (‘When I don’t know if I should respond to an email, I do it anyway’), real time emailing (‘When I have question, I email it directly’), and email addressee uncertainty (‘When I don’t know who to email, I email everyone’). The three proposed self-nudges are (1) providing your colleagues a timely reminder about whether they need to respond to an email or not (hereby also indirectly reminding yourself of this behavior), (2) providing reminders to yourself about the question you have so you can delay action and have time to think about alternative strategies, (3) reframing emailing to be about the receiver rather than the sender: considering the consequences of limiting the number of addressees for colleagues (reduced stress) rather than for yourself (it may take a bit more time before you receive the right answer).

*Image 3: Self-nudges, image with behavioral challenges.*

*Image 4: Self-nudges, self-nudge 1.*

*Image 5: Self-nudges, self-nudge 2.*

*Image 6: Self-nudges, self-nudge 3.*

**Appendix C: Survey measures**

This appendix introduces the survey measures, including an elaborate explanation of the Bayesian Truth Serum. In the survey experiment, questions were translated to Dutch.

***Email volume (Sumecki et al., 2011)***

Measured with two open questions with numerical content validation:

*You’ve been on holiday for one week. How many new emails would you expect to find in your inbox when you return to work? (in digits)*

*On an average working day, how much time (in minutes) do you spend managing emails (reading, sending, filing, etc.)? (in digits)*

***Email overload (Dabbish and Kraut, 2006)***

Measured with seven items on a 7-point Likert scale ranging from ‘strongly disagree’ to ‘strongly agree’:

1: *I can handle my email efficiently.* (Reversed)

2: *I have trouble finding information in my email.*

3: *I can easily deal with the amount of email I receive.* (Reversed)

4: *I sometimes miss information or important email messages.*

5: *I reply quickly to the email message I need to*. (Reversed)

6: *Dealing with my email disrupts my ongoing work.*

7: *I find dealing with my email overwhelming.*

***Feasibility, appropriateness, meaningfulness and effectiveness (FAME-approach for evidence-based practice, Jordan et al., 2019)***

Measured with multiple separate 7-point Likert scales:

*How feasible would it be to use this message in your organization?* (‘Very unfeasible’ to ‘Very feasible’)

*How appropriate would this message be in your organization?* (‘Very inappropriate’ to ‘Very appropriate’)

*How meaningful would this message be to your organization?* (‘Very meaningless’ to ‘Very meaningful’)

*How effective would this message be in decreasing email use in your organization?* (‘Very ineffective’ to ‘Very effective’)

***Bayesian Truth Serum for non-compliance (John et al., 2012; Prelec, 2004)***

The Bayesian truth serum is a scoring algorithm that combines the answers of respondents about their behavior and their estimates of what others would answer (John et al., 2012, p. 526). The serum increases credibility by developing multiple estimates (elaborated below) rather than one estimate about respondents’ own behavior (i.e., would you comply) and combining these estimates into a conservative judgement of, in our case, non-compliance. The serum has been used in large-scale surveys (Van de Schoot et al., 2021; Frank et al., 2017; Weaver and Prelec, 2013) and scholars recommend using this approach in experimental social science research (Schoenegger, 2023).

In our setup, it worked as follows. First, we briefly explained its purpose to respondents: ‘we would like to ask you to predict the response of your colleagues and yourself. The following questions help us predict the effect of this message’. We then asked respondents to estimate three values: 1) the percentage of colleagues that would send as many emails after the intervention as before (the prevalence estimate), 2) the percentage of colleagues that would be honest about sending as many emails after the intervention as before (the admission estimate), and 3) whether they would send less emails after the intervention (yes or no; the self-admission rate). This method allows us to report three different estimates of non-compliance: self-admission rates, prevalence estimates, and prevalence estimates calculated by dividing the self-admission rates by the admission estimates. We generate a more conservative judgement of non-compliance by taking the geometric mean of these three values (John et al., 2012) as indicated in formula 1. Subtracting the proportion of non-compliance from 100 gives us the proportion of compliance. In our results, we will compare the self-admission estimates to the geometric mean.

(1)

The specific questions are presented below:

Introduction: *We would like to ask you to predict the response of your colleagues and yourself. The following questions help us predict the effect of this message.*

1: *What percentage of your colleagues would send as many emails after this message as before?* (from 0% to 100% using a graphic slider) (prevalence estimate)

2: *What percentage of your colleagues would be honest about sending as many emails after this message as before?* (from 0% to 100% using a graphic slider) (admission estimate)

3: *Are you going to send less emails after this message?* (Yes/No) (self-admission rate)

***Perceived autonomy (Morgeson and Humphrey, 2006; Gorgievski et al., 2016)***

Measured with three items on a 7-point Likert scale ranging from ‘strongly disagree’ to ‘strongly agree’:

1: *This message gives me a chance to use my personal judgement in using email.*

2: *This message allows me to make a lot of decisions about using email on my own.*

3: *This message provides me with significant autonomy in making decisions about using email.*

***Work engagement (UWES-3, Schaufeli et al., 2019)***

Measured with a 5-point Likert scale ranging from ‘Never’ (1) to ‘Always (daily)’ (5):

1: *At my work, I feel bursting with energy.*

2: *I am enthusiastic about my job.*

3: *I am immersed in my work.*

**Appendix D: Methods and results pilot study**

This appendix describes the methods and results of the pilot study.

*Methods* Participants

We collected data on March 22 from 435 respondents via Prolific in an English language Qualtrics survey. The sample size was determined with an a priori power analysis using G\*Power. The survey did not inquire about personal data except those provided by Prolific itself, for which respondents provided consent. Respondents were required to work full-time and use email in their job. The mean age of the 435 respondents was 32.75 (SD = 8.71, Min. = 19, Max. = 69). Regarding gender, 204 were female, 230 were male and 1 respondent indicated they would rather not say.

Procedure and measures

Respondents first provided background characteristics, including an eligibility check (respondents had to work full-time and use email at their job). We assessed email volume and email time with open questions adapted from Sumecki et al., 2011 (p. 409): ‘You’ve been on holiday for one week. How many new emails would you expect to find in your inbox when you return to work? (in digits)’ and ‘On an average working day, how much time (in minutes) do you spend managing emails (reading, sending, filing, etc.)? (in digits)’. We also assessed email overload with 7 items on a 7-point Likert scale (α = .82) ranging from ‘strongly disagree’ to ‘strongly agree’ (Dabbish and Kraut, 2006). Next, respondents were exposed to one of the three nudges randomly. Chi-square tests indicated that randomization was successful among gender (male versus female) and age groups (younger versus older than mean age) as no significant differences existed (χ2(2) = .42, p = .813 for gender and χ2(2) = .13, p = .936 for age). The instruction accompanying the nudges read ‘Imagine the organization you work for sends you the following message about using email in your organization. (If you fill out this survey on a mobile phone, you are able to zoom in.) Please read the message carefully.’ After the nudge, they were asked to assess perceived autonomy on a 7-point Likert scale (α = .86), a scale we created by adapting the three item Decision-Making Autonomy subscale from the Work Design Questionnaire (WDQ; Morgeson and Humphrey, 2006). Inspired by Jordan et al. (2019), we also assessed the feasibility, appropriateness, meaningfulness and expected effectiveness of the nudge with single items on 7-point Likert scales, asking respondents to imagine the effects of this message in their organization.

To come to our final sample (N = 435), listwise deletion was applied and respondents had to pass an attention check (a multiple choice question with two boxes and the instruction ‘please only check box 2’). Three respondents failed this check. We use one way analysis of variance for our main analysis and Kruskal-Wallis H Tests for additional analyses (these were used as we used ordinal variables. For interpretation, means and standard deviations were included). Significance levels were set at p = .05 (for all models, exact p-levels were reported).

*Results*  
Table D1 present the means, standard deviations and correlations of the main variables.

*Table D1: correlations. N = 435.*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **M (SD)** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| 1. Email volume | 106.88 (257.23) | - | - | - | - | - | - | **-** | **-** |
| 2. Email time | 73.24 (78.52) | .35\* | - | - | - | - | - | **-** | **-** |
| 3. Email overload | 2.64 (1.03) | .26\* | .20\* | - | - | - | - | **-** | **-** |
| 4. Autonomy | 5.26 (1.18) | .03 | .03 | -.06 | - | - | - | **-** | **-** |
| 5. Feasibility | 4.56 (1.65) | .003 | -.001 | .02 | .31\* | - | - | **-** | **-** |
| 6. Appropriateness | 4.86 (1.50) | .07 | .07 | .06 | .41\* | .67\* | - | **-** | **-** |
| 7. Meaningfulness | 4.58 (1.54) | .08 | .05 | .07 | .47\* | .67\* | .69\* | **-** | **-** |
| 8. Effectiveness | 4.11 (1.58) | -.03 | .00 | .008 | .43\* | .56\* | .54\* | .71\* | **-** |

*\* p < .001 (two-tailed).*

*Note*

*Correlations are Pearson except for those with email volume and email time, these are Spearman as for these variables the data indicated outliers.*

A one way analysis of variance showed that the effects of the nudges on perceived autonomy differed significantly, F(2,432) = 5.21, p = .006 (η2 = .024)[[1]](#footnote-2). Tukey HSD post hoc analyses indicated that the perceived autonomy was significantly lower for the rule-of-thumb (M = 5.01, SD = 1.26) compared to the opinion leader nudge (M = 5.41, SD = 1.15) (p = .009) and the self-nudges (M = 5.37, SD = 1.07) (p = .026). Perceived autonomy did not differ significantly across age, gender or email overload (t-test with dummy variable younger/older than the mean age: t(433) = .42, p = .672[[2]](#footnote-3); t-test with dummy variable male/female: t(432) = -1.77, p = .078[[3]](#footnote-4); t-test with dummy variable below mean email overload/above mean email overload: t(431.46) = -.39, p = .699[[4]](#footnote-5)).

We conducted Kruskal-Wallis H Tests and found that effects of the nudges differed on perceived feasibility, appropriateness and meaningfulness, but not effectiveness (table D2). Pairwise comparisons indicate that for feasibility, the opinion leader nudge scored significantly (p < .05) lower than the other nudges. For appropriateness, the self-nudges scored significantly higher than the other nudges. For meaningfulness, the self-nudges scored significantly higher than the opinion leader nudge.

*Table D2: Kruskal-Wallis H Tests and descriptives.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Feasibility** | **Appropriateness** | **Meaningfulness** | **Effectiveness** |
| *Test result* | H(2) = 19.49,  p < .001 | H(2) = 17.62,  p < .001 | H(2) = 6.32,  p = .042 | H(2) = 4.35, p = .114 |
| ***Means (SDs) and pairwise comparisons*** | | | | |
| *Opinion leader* | 4.11 (1.69)†‡ | 4.55 (1.55)† | 4.38 (1.62)† | 4.03 (1.63) |
| *Rule-of- thumb* | 4.63 (1.63)† | 4.75 (1.51)‡ | 4.53 (1.52) | 3.97 (1.65) |
| *Self-nudges* | 4.94 (1.52)‡ | 5.28 (1.35)†‡ | 4.83 (1.47)† | 4.35 (1.45) |

*† and ‡: Categories that have significantly differing mean rank scores cf. the Kruskal-Wallis H Tests (p < .05). Significance values were adjusted with the Bonferroni correction for multiple tests.*

**Appendix E: Traditional email interventions.**

The traditional interventions were based on real strategies that organizations have to reduce email stress, resembling traditional policy instruments (Tummers, 2019). The first intervention is technological. Organizations could limit email access to specific hours in a day so that employees are limited to checking email a few times a day. We term this the ‘email access limit’. It resembles the ‘whip’ approach (Tummers, 2019) as well as organizational strategies to force employees to take time off. The next intervention is an economic incentive and resembles the ‘carrot’ approach (Tummers, 2019). Organizations could reward employees financially for reducing their email use. We term this the ‘monetary reward’. Offering employees monetary rewards for meeting goals is a very common business practice (Aguinis et al., 2013). The last intervention is variation on the ‘carrot’ by means of a social incentive. In the intervention ‘public praise’, employees are publicly praised for showing exemplary behavior in email use. Research suggests public praise may be more effective than financial incentives (Handgraaf et al., 2013).

*Table E1: traditional interventions.*

|  |  |  |
| --- | --- | --- |
| **Intervention** | **Dutch text** | **English translation** |
| Email access limit | Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom kunnen medewerkers vanaf nu alleen tussen 10 en 11 uur ’s ochtends en 3 en 4 uur ‘s middags e-mail versturen of ontvangen. We hopen dat dit helpt om je een legere mailbox te bezorgen | There is a lot of emailing within our organization. This can cause stress. Therefore, from now on, employees can only send or receive e-mail between 10 a.m. and 11 a.m. and 3 p.m. and 4 p.m. We hope this helps to give you an empty mailbox. |
| Monetary reward | Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom hebben we uitgerekend hoeveel e-mails jij gemiddeld per werkdag verstuurt. Vanaf nu krijg je per werkdag voor elke e-mail die je minder stuurt dan dit gemiddelde, 1 euro extra bij je volgende salarisstrook. We hopen dat dit helpt om je een legere mailbox te bezorgen. | There is a lot of emailing within our organization. This can cause stress. That is why we have calculated how many e-mails you send on average per working day. From now on, per working day you will receive 1 euro extra for every e-mail that you send less than this average. We hope this helps to give you an empty mailbox. |
| Public praise | Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom sturen we vanaf nu elke week een lijst met ‘minder e-mail helden’ rond: dit zijn de medewerkers die binnen hun functie de minste e-mails hebben verstuurd. We hopen dat dit helpt om je een legere mailbox te bezorgen. | There is a lot of emailing within our organization. This can cause stress. That is why from now on we will send out a list of 'less-email heroes' every week: these are the employees who have sent the fewest emails within their position. We hope this helps to give you an empty mailbox. |

**Appendix F: Randomization survey experiment**

To assess whether randomization between the different interventions in the survey experiment was successful, we computed multiple chi-square tests for independence.

**Gender**

Table F1 presents the division of gender across interventions. A Chi-square test indicated that randomization was successful among gender (male or female): χ2(6) = 2.48, p = 0.871.

*Table F1: Gender across interventions.\**

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervention** | **Gender** |  | ***Total*** |
|  | **Female** | **Male** |  |
| **Opinion leader** | 510 | 73 | 583 |
| **Rule-of-thumb** | 509 | 83 | 592 |
| **Self-nudges** | 491 | 88 | 578 |
| **All nudges** | 493 | 84 | 577 |
| **Email access limit** | 506 | 83 | 589 |
| **Monetary reward** | 496 | 89 | 585 |
| **Public praise** | 501 | 87 | 588 |
| ***Total*** | 3506 | 587 | 4093 |

\*Respondents reporting X or that they would rather not say (N = 19) were left out of this analysis.

**Age groups**

Table F2 presents the division of age groups across interventions. A Chi-square test indicated that randomization was successful among age groups: χ2(30) = 32.29, p = 0.354.

*Table F2: Age groups across interventions.\**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Intervention** | **Age groups** | |  |  |  |  | ***Total*** |
|  | **25 or younger** | **26-35** | **36-45** | **46-55** | **56-65** | **66 or older** |  |
| **Opinion leader** | 6 | 46 | 102 | 169 | 258 | 7 | 588 |
| **Rule-of-thumb** | 3 | 46 | 90 | 180 | 265 | 7 | 591 |
| **Self-nudges** | 6 | 35 | 97 | 186 | 252 | 8 | 584 |
| **All nudges** | 7 | 30 | 89 | 182 | 261 | 9 | 578 |
| **Email access limit** | 3 | 34 | 116 | 195 | 238 | 4 | 590 |
| **Monetary reward** | 3 | 34 | 131 | 164 | 248 | 7 | 587 |
| **Public praise** | 3 | 34 | 100 | 190 | 255 | 9 | 591 |
| ***Total*** | 31 | 259 | 725 | 1266 | 1777 | 51 | 4109 |

\*A total of 3 respondents did not disclose age.

**Healthcare sector**

Table F3 presents the division of healthcare sectors across interventions. A Chi-square test indicated that randomization was successful among healthcare sector: χ2(24) = 24.25, p = 0.447.

*Table F3: Healthcare sectors across interventions.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Intervention** | **Healthcare sectors** | |  |  |  | ***Total*** |
|  | **Hospitals** | **Nursing/home care** | **Mental healthcare** | **Disabled care** | **Other** |  |
| **Opinion leader** | 221 | 156 | 101 | 79 | 32 | 589 |
| **Rule-of-thumb** | 226 | 156 | 74 | 91 | 45 | 592 |
| **Self-nudges** | 213 | 142 | 99 | 97 | 33 | 584 |
| **All nudges** | 231 | 140 | 92 | 85 | 30 | 578 |
| **Email access limit** | 199 | 152 | 98 | 90 | 52 | 591 |
| **Monetary reward** | 206 | 155 | 91 | 95 | 40 | 587 |
| **Public praise** | 219 | 158 | 98 | 83 | 33 | 591 |
| ***Total*** | 1515 | 1059 | 653 | 620 | 265 | 4112 |

**Working hours**

Table F4 presents the division of working hours across interventions A Chi-square test indicated that randomization was successful among amount of working hours per week (29 hours or more versus less than 29): χ2(6) = 2.589, p = 0.858.

*Table F4: Working hours across interventions.\**

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervention** | **Working hours** |  |  |
|  | **Less than 29** | **29 or more** | ***Total*** |
| **Opinion leader** | 282 | 304 | 586 |
| **Rule-of-thumb** | 281 | 309 | 590 |
| **Self-nudges** | 281 | 303 | 584 |
| **All nudges** | 270 | 306 | 576 |
| **Email access limit** | 300 | 290 | 590 |
| **Monetary reward** | 278 | 307 | 585 |
| **Public praise** | 291 | 297 | 588 |
| ***Total*** | 1983 | 2116 | 4099 |

\*Respondents who reported to have a zero-hours contract (N = 13) were left out of this analysis.

**Appendix G: Preregistration and evaluation of original hypotheses**

**Preregistration and deviations**

Initially, the pilot study, the survey experiment and the quasi-field experiment were all preregistered separately with a total of four hypotheses (Van Roekel et al., 2022a; 2022b; 2022c). In this paper, we combined the initial preregistrations into one main hypothesis. Below we explain how, and we briefly discuss all original hypotheses.

Specifically, we moved the pilot study to an appendix. The hypothesis of the pilot study focused on the slight differences in perceived autonomy between the nudges but we decided that the main paper should rather focus on the autonomy and effectiveness of nudges in general. Second, to come to our one main hypothesis, we merged all three remaining hypotheses from the survey experiment and quasi-field experiment. The original hypotheses separately addressed nudges 1) preserving autonomy more than traditional interventions, 2) being perceived as less effective than traditional interventions, 3) being effective in decreasing email use. We decided to introduce the traditional interventions in the method section and not explicitly mention them in the theory section. Therefore, our main hypothesis in the paper focuses on the expectation that nudges are autonomy-preserving and effective in decreasing email use. In our method section, we now describe how we test this by evaluating the absolute and relative scores (the latter in comparison to traditional interventions) of autonomy and nudge effectiveness, and by testing both subjective and objective nudge effectiveness.

Regarding the preregistration of analyses, one major deviation should be mentioned. The quasi-field experiment was analyzed with a slightly different statistical analysis, linear mixed models rather than ANOVA. The preregistered analyses appeared less suitable for analysis when the data came in, and linear mixed models allowed for more flexibility. Specifically, this analysis allowed for fixed factors to be included, in our analysis time (the week) was the repeated measure fixed factor (Krueger and Tian, 2004). The change of analysis did not fundamentally change our results.

**Evaluation of original hypotheses**

*Pilot study*

We expected self-nudges to be most autonomy-preserving, as it combined full transparency with the ability to influence one’s own choice environment. We expected an opinion leader nudge to be the least autonomy-preserving, as it employs hierarchy (i.e., the opinion leader) to make employees change behavior. The rule-of-thumb was estimated to score in between the self-nudges and the opinion leader nudge, as it does not employ hierarchy but does suggest a specific behavior change.

H1: *A rule-of-thumb nudge will preserve autonomy more than an opinion leader nudge (a) but less than self-nudges (b).*

The results in appendix C show that perceived autonomy was significantly lower for the rule-of-thumb in comparison to the other nudges. Therefore, the first part (a) of the above hypothesis was not confirmed, but the second part (b) was. A possible explanation to why the rule-of-thumb was considered the least autonomy-preserving, may be that this nudge was more specific in telling employees how to behave. What is more, employees may not have experienced the opinion leader nudge as imposing hierarchical pressure. The implicit hypothesis underlying this hypothesis, namely that nudges will preserve autonomy, is evaluated in the main paper.

*Survey study*

Two hypotheses were preregistered for the survey study. They both specifically compare nudges to the traditional interventions. In the final paper, we decided to move the introduction of the traditional interventions to the method section, and have it be only part of the evaluation of the hypothesis besides evaluating the absolute scores on perceived autonomy and effectiveness. Nevertheless, the first hypothesis below was confirmed by our results: nudges preserve autonomy more than traditional interventions. The second hypothesis, however, was not: we expected employees may predict that the traditional interventions would be more effective as they offered more rigorous (yet less autonomy-preserving) ways to change behavior. In contrast to this expectation, employees thought the nudges would be more effective.

H1: Nudges will preserve autonomy more than technical, social or economic interventions.

H2: Employees will predict that technical, social or economic interventions are more effective than nudges.

*Quasi-experiment*

The original hypothesis for the quasi-experiment stated nudges would be effective in decreasing actual email use, and is still evaluated as part of the combined hypothesis in the paper. While we observe decreases in email use, the partially insignificant results suggest we should be careful in our conclusions.

H1: An opinion leadership nudge, rule-of-thumb and self-nudges will decrease email volume.

1. Equal variances assumed as F(2,432) = 2.06, p = .129. [↑](#footnote-ref-2)
2. Two-sided p. Equal variances assumed as F(2,432) = 2.81, p = .094. [↑](#footnote-ref-3)
3. Two-sided p. Equal variances assumed as F(2,432) = .31, p = .577. [↑](#footnote-ref-4)
4. Two-sided p. Equal variances not assumed as F(2,432) = 5.08, p = .025. [↑](#footnote-ref-5)