**Experiment 1**

The main goal of the experiments is to test whether people infer not only that convergent estimates are more likely to be correct, but also that, as a result, people whose estimates converge are more likely to be competent, in particular if the convergence is precise. In the following, we use the term “informants” for those individuals who provide information to the participants, and “estimates” for the information they provide. We define “convergent” estimates simply as numeric guesses that are close to each other.

For the convergence of estimates to be a reliable cue to the accuracy of the estimates, even in the absence of priors about the competence of informants, the estimates have to be independent (and they must reflect the best estimates of the players). Here, we manipulate the degree to which numerical estimates converge. We do so across two numbers of estimates (3 and 10) to establish some potential robustness for our findings (i.e. that they are not only true for a one number of estimates).

Based on the literature on information aggregation (e.g. Mercier & Morin, 2019), we form a first hypothesis:

> H1: *When making a guess based on the estimates of (independent) informants, participants will be more confident about their guess when these estimates converge compared to when they don’t.*

Our second, and main hypothesis is:

> H2: *Participants perceive (independent) informants whose estimates converge more as more competent than informants whose estimates converge less.*

We will also investigate the following research questions:

> RQ1: *Do H1 and H2 hold for both a small [3] and a large [10] number of estimates?*
>
> RQ2: *When making a guess based on the opinions of (independent) informants, will participants be more confident about their guess when there is a larger number of estimates compared to when this number is smaller?*
>
> RQ3: *Is there an interaction effect between the number of estimates and convergence on perceived competence of informants?*

**Materials.** “Some people are playing games in which they have to estimate various quantities. Each game is different. You have no idea how competent the people are: they might be completely at chance, or be very good at the task. It’s also possible that some are really good while others are really bad. Some tasks might be hard while others are easy. Across various games, we will give you the answers of several players, and ask you questions about how good they are.

As it so happens, for all the games, the estimates have to be between 1000 and 2000, but all the games are completely different otherwise, and might require different skills, be of different difficulties, etc.

Each player in the game makes their own estimate, completely independent of the others".

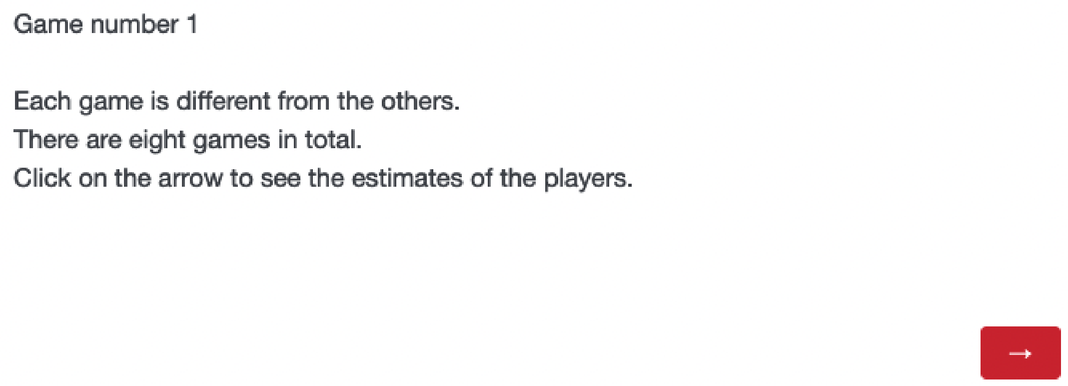



***Fig.1:*** *Reminder before each game*

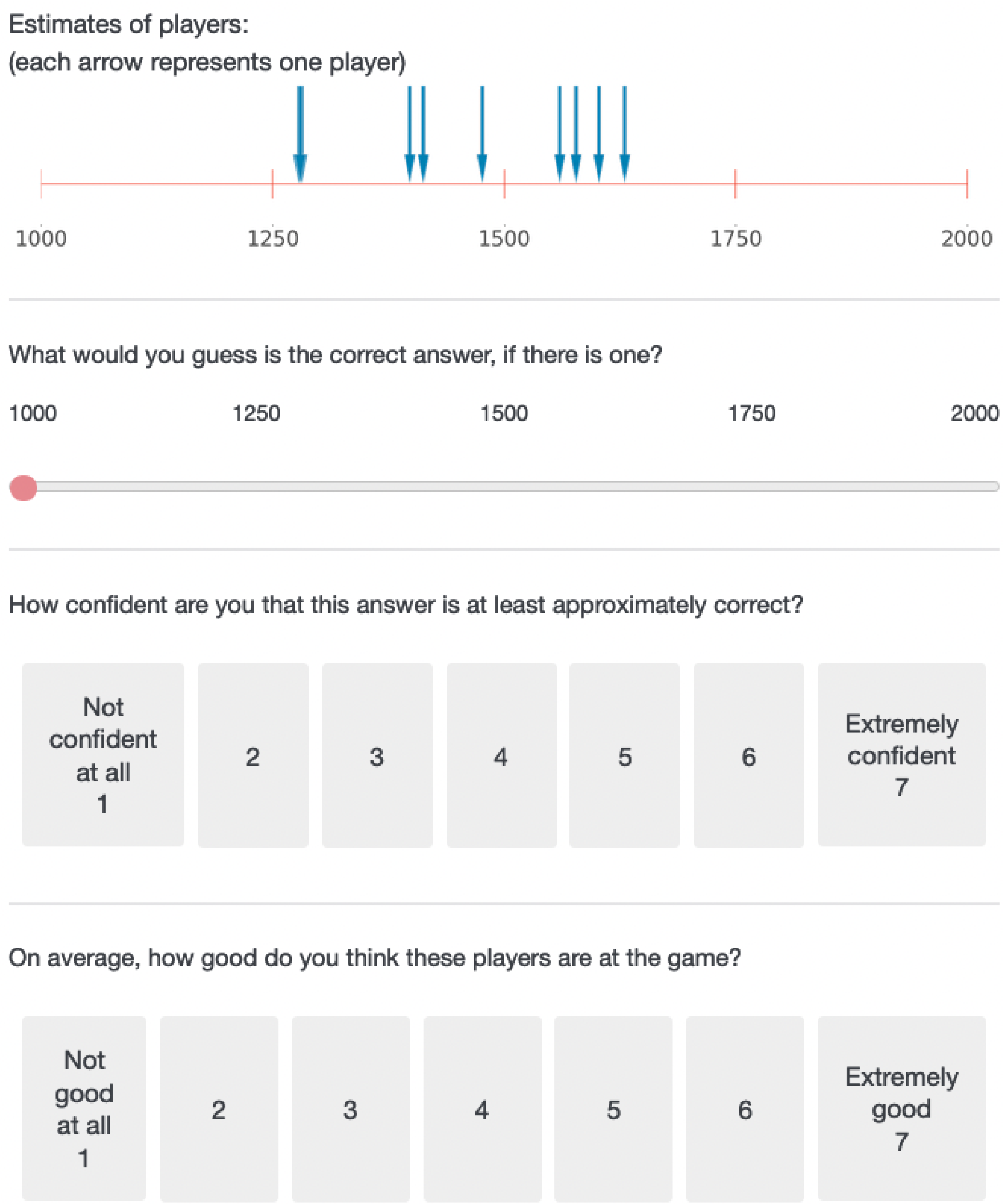



***Fig.2:*** *Results of one of 8 games that participants have to rate.*

**Design and procedure.** We manipulate two experimental factors: (i) the convergence of the estimates (how close they are) and (ii) the number of estimates (how many players there are). Each factor has two levels, low and high. We use a 2(convergence: low/high) x 2(number: low/high) design. We use a within-participant design, with each participant seeing all of the conditions. Participants will see two sets of estimates (game results) per condition. Thus, each participant will see eight sets of estimates. Participants are told that each game is different, with a different solution. The order in which individual participants see sets of estimates will be randomized. There is a total of three series of eight sets of estimates, and each participant sees one series.

**Stimuli.** The sets of estimates provided to the participants are generated with random draws from normal distributions. First, we vary the standard deviation of these distributions to simulate the degree of convergence (20 for low convergence, 150 for high convergence). Second, we vary the number of draws from these distributions (three for low number, 10 for high number). For each condition, we take two random draws from the respective normal distribution.

The means of the normal distributions that we draw our estimates from are distinct between sets of estimates. Considering our within-participant design, this makes it more likely that participants understand each set of estimates as being the result of a different game, with a different solution. In order to assure that random draws from the distributions will (most likely) appear on the response scale (ranging from 1000 to 2000), we constrain the means of all normal distributions to lie between the first and third quartile of the response scale (i.e. smallest mean possible is 1225 and largest is 1775). We define a set of eight means–one for each set of estimates–that cover the range from first to third quartile of the predefined scale with an equal interval (1250, 1325, 1400, 1475, 1550, 1625, 1700, 1775).

**Randomization.** We randomly pair means with conditions. We then draw the set of estimates from the respective normal distributions given the assigned means and the condition constraints. We repeat this three times, resulting in three series of eight sets of estimates. We randomly assign participants to one of these series. Additionally, for each participant, we randomize the order of appearance of the sets of estimates within the respective series.

**Dependent variables.** For each set of estimates participants respond to the following questions: First, we ask participants to make a guess about the correct answer based on the estimates they see ("What would you guess is the correct answer, if there is one?"). Participants indicate their numeric guess using a slider on a line identical with the one they see the estimates on. Participants are then asked about how confident they are regarding their own guess ("How confident are you that your answer is at least approximately correct?" on a 7-point Likert scale ("not confident at all" to "extremely confident"). Finally, participants are asked how they perceive the competence of the group of players whose estimates they saw in a game. Competence ("On average, how good do you think these players are at the game?") is assessed on a 7-point Likert scale (from "not good at all" to "extremely good").

**Statistical analysis.** All analyses will be conducted in R (v.4.1.1) using R Studio.

H1: We will perform a paired t-test to assess the effect of *convergence* on participants' *confidence* about their guesses. This test accounts for the dependency between the two conditions of convergence issued by our within-participant design. [R code: *t.test(data$confidence, data$convergence, paired = T)*]

H2: We will perform a paired t-test to assess the effect of *convergence* on participants' *competence* ratings. [R code: *t.test(data$competence, data$convergence, paired = T)*]

For a robustness check, we will run a mixed model regression analysis of *convergence* on *competence* ratings. The model will include random intercepts for participants and control for a fixed effect of our second experimental factor, the *number* of estimates. Including the factor *number* as a control will also address RQ1. [R code: *lmer(competence ~ convergence + number + (1 | participant), data)*]

RQ2: We will perform a paired t-test to assess the effect of *number* on participants' *confidence* about their guesses. [R code: *t.test(data$confidence, data$number, paired = T)*]

RQ3: We will run a mixed model regression with an interaction of the fixed effects of *convergence* and *number* on *competence*. The model will include random intercepts for participants. [R code: *lmer(competence ~ convergence*number + (1 | participant), data)*]

**Exclusions.** We will exclude participants failing (i.e. participants not answering the question or writing "yes" or "no" instead of "I pay attention") the following attention check:

*Imagine you are playing video games with a friend and at some point your friend says:*

*"I don't want to play this game anymore! To make sure that you read the instructions, please write the three following words "I pay attention" in the box below. I really dislike this game, it's the most overrated game ever."*

*Do you agree with your friend?*

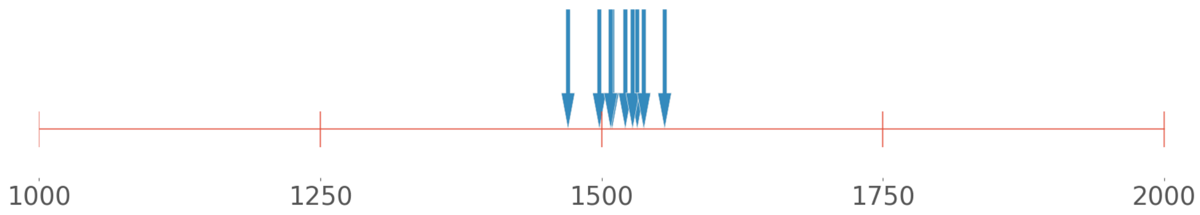
**Power analysis.** We performed an a priori power analysis with G*Power3. To compute the necessary number of participants, we decided that the minimal effect size of interest would correspond to a Cohen's d of 0.2 between two different experimental conditions, since this corresponds to what is generally seen as a small effect. We chose a t-test to detect the difference between two dependent means (matched pairs). This test is taking into account our within-participant design. In order to detect an effect size of Cohen's d = 0.2 with 80% power (alpha = .05, two-tailed), G*Power suggests we would need 199 participants in a paired samples t-test.

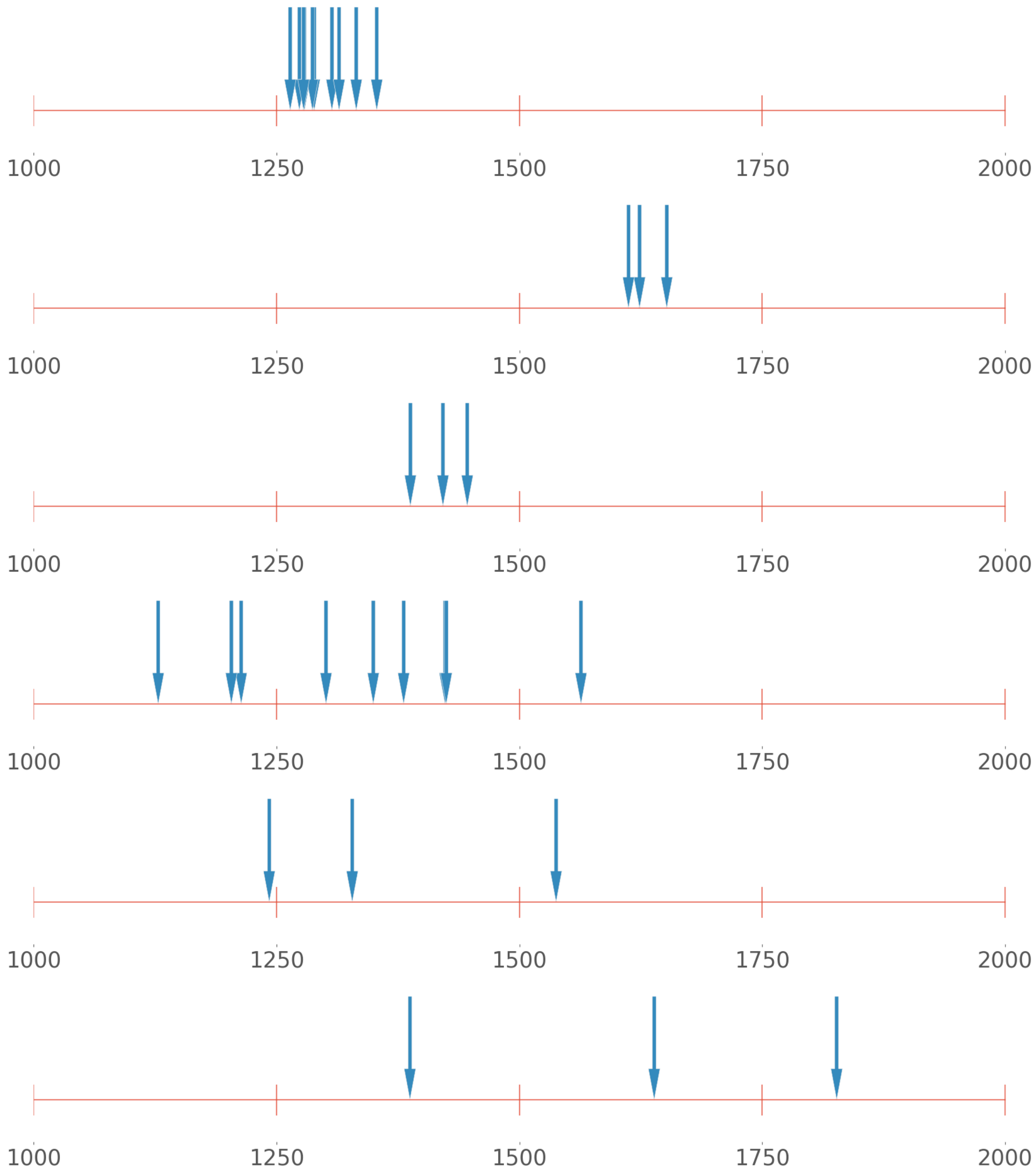
**References**

Mercier, H., & Morin, O. (2019). Majority rules: how good are we at aggregating convergent opinions?. Evolutionary Human Sciences

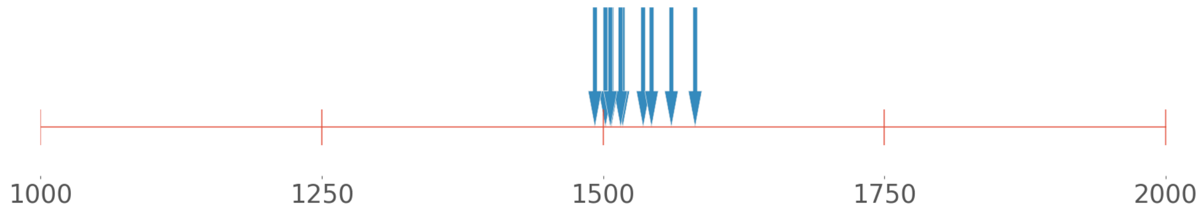
**Appendix**

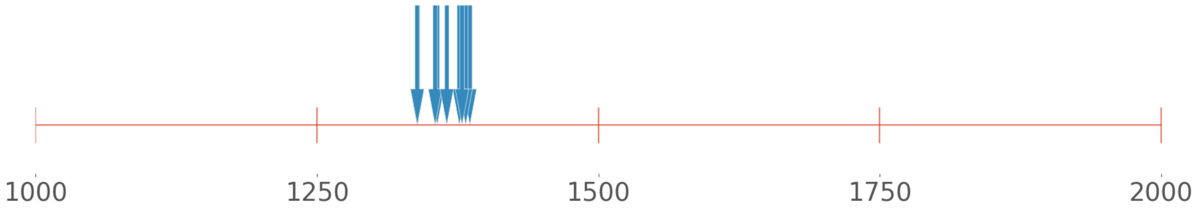
a) Sets of estimates
   Series 1:

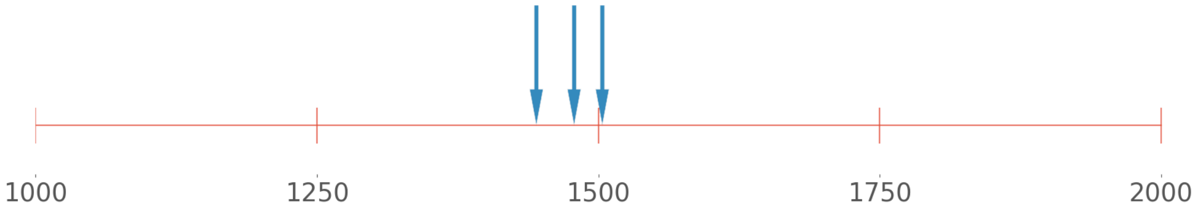


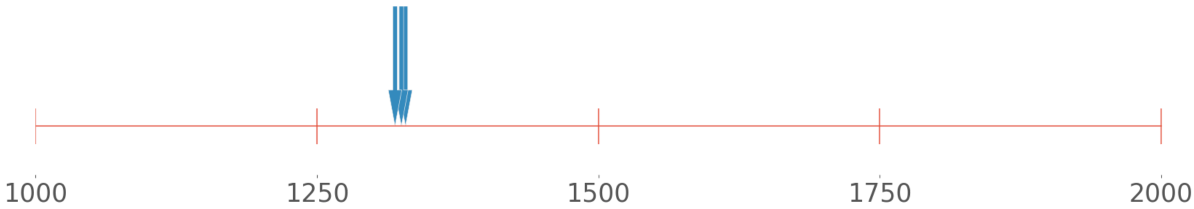


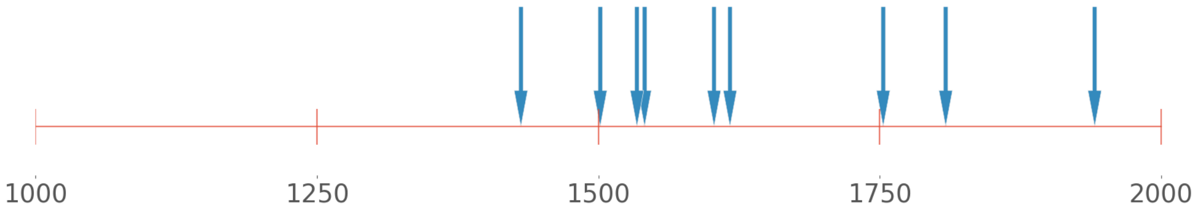

Series 2:

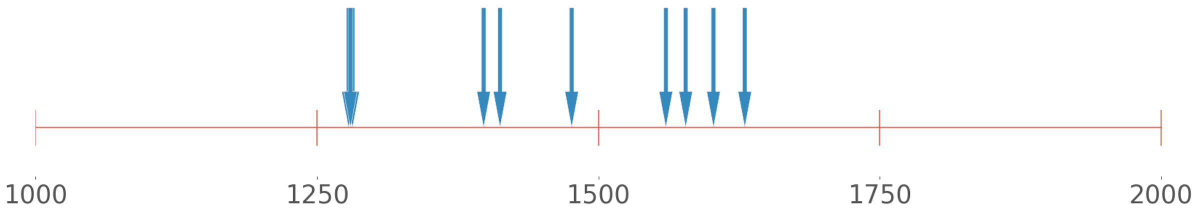


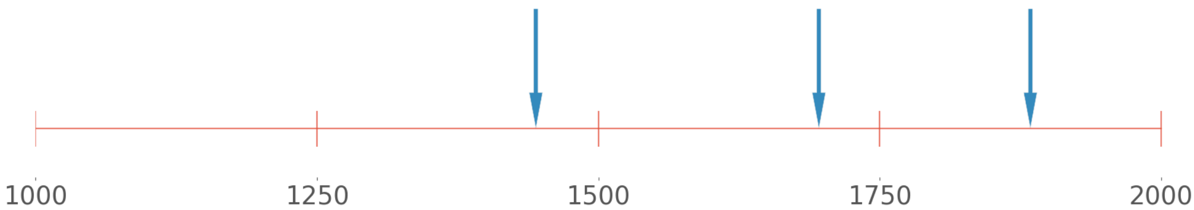

1000
1250
1500
1750
2000

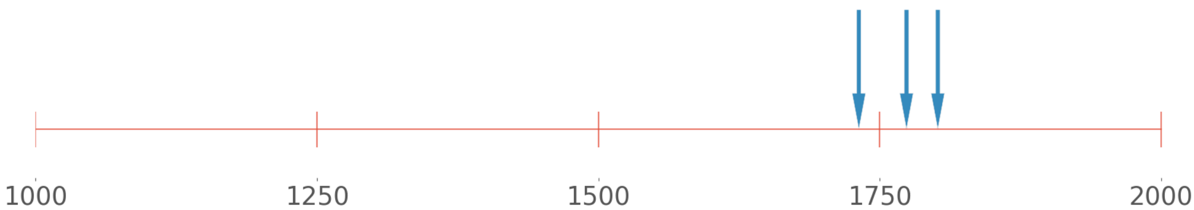

1000
1250
1500
1750
2000


1000
1250
1500
1750
2000

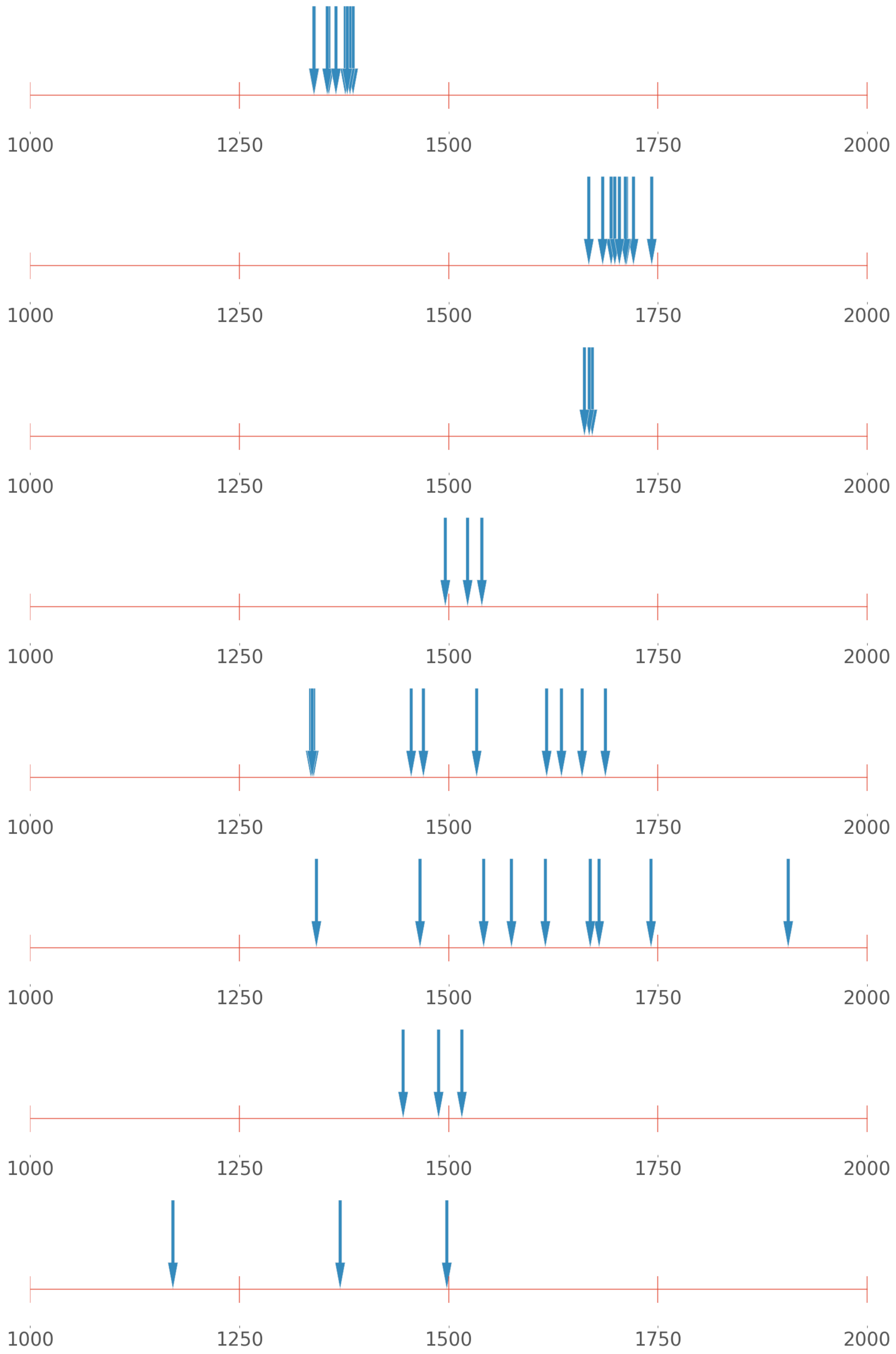

1000
1250
1500
1750
2000


1000
1250
1500
1750
2000


1000
1250
1500
1750
2000


1000
1250
1500
1750
2000

Series 3:


1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000
1000 1250 1500 1750 2000