Experiment 3: Convergence and Independence

In experiment one, we argued that 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*. In experiment three, we test this independence assumption. Our main research question is: Does the effect of convergence on competence depend on whether informants are independent or not?

In contrast to experiment 1, we use a less abstract setting: participants see the fictional predictions of three experts for stock values. We manipulate two factors: *independence* (with the levels independence and conflict of interest) and *convergence* (with the two levels convergence and divergence).

We will manipulate *convergence* within participants. That is, all participants will see four sets of predictions: two in which the predictions of the experts converge and two in which they diverge. We will manipulate *independence* between participants. That is, each participant will be randomly assigned to only one of the following conditions: In the independent condition, participants will be presented with a scenario in which experts made their predictions independently–that is, on their own and without a conflict of interest. In the conflict of interest condition, experts personally gain from making predictions close to a certain value.

We expect the findings of experiment one to replicate in the new setting of experiment two. In this experiment, independence of informants was given, and convergence or divergence of their estimates manipulated. Therefore, our first set of hypotheses is:

> H1a: Participants perceive predictions of **independent** informants as **more accurate** when they converge compared to when they don't.
>
> H1b: Participants perceive **independent** informants as **more competent** when their predictions converge compared to when they don't.

How about a context in which informants are not independent but in a conflict of interest instead? If we are right in assuming that the positive effect of convergence on competence hinges upon the fact that informants are independent, we should observe a different, smaller effect in a context where they are not.

> H2a: The e**ffect of convergence on accuracy (H1a)** is **more positive** in a context where informants are **independent** compared to when they are in a conflict of interest.
>
> H2b: The e**ffect of convergence on competence (H1b)** is **more positive** in a context where informants are **independent** compared to when they are in a conflict of interest.

In H2, we do not specify what form the effect of convergence on perceived competence takes on when informants are in a conflict of interest. All we predict is that it is less positive – it could be positive, negative or even non-existent. Our research questions further inquire about this effect. Although we do not make predictions about it, we believe that informants could in fact be less trusted and deemed less competent when their predictions converge in an answer that suits their personal interest. That is because, by contrast, if informants with those same incentives do *not* converge, this suggests that at least some of them provide non-incentivised, hence more likely truthful predictions.

> RQ1: Do participants perceive predictions of informants **in a conflict of interest** as **less accurate** when they converge compared to when they don't?
>
> RQ2: Do participants perceive **informants in a conflict of interest** as **less competent** when their predictions converge compared to when they don't?

Materials

Introduction. "You will see four scenarios in which several experts predict the future value of a stock. You have no idea how competent the experts are. It's also possible that some are really good while others are really bad.

As it so happens, in all scenarios, the predictions for the value of the stock have to lie between 1000 and 2000. Other than that, the scenarios are **completely unrelated**: it is different experts predicting the values of different stocks every time."

[page break]

Introducing the first scenario as: "In the first scenario, [condition]".

Introducing the other scenarios as: "In scenario number N, three different experts predict the value of a different stock.

[Independence] Otherwise, circumstances are the same: [condition]

[Conflict of interest] Again, all three experts have invested in the specific stock whose value they are predicting. This time, these experts benefit if other people believe that the stock will be valued at *[mean of distribution]* in the future. "

Independent condition. Experts are independent of each other, and have no conflict of interest in predicting the stock value - they do not personally profit in any way from any future valuation of the stock.

Conflict of interest condition. All three experts have invested in the specific stock whose value they are predicting, and they benefit if other people believe that the stock will be valued at *[mean of distribution]* in the future.

Design and procedure

We manipulate the factor *independence* (two levels, *independence* and *conflict of interest*) between participants and the factor convergence (two levels, *convergence* and *divergence*) within. By contrast to experiment one, participants will see only two sets of estimates (stock predictions) per *convergence* condition (i.e four in total). The number of experts making predictions in each set of estimates is always the same: three. The order in which individual participants see the four sets will be randomized.

Stimuli. Our design requires that each participant sees four different sets of predictions, two convergent and two divergent ones, that appear on a scale from 1000 to 2000. For all four sets, the number of predictions is always the same: three. In contrast with experiment one, the sets of estimates provided to the participants are generated with random draws from *uniform* distributions (instead of *normal* distributions in the previous experiment). We vary the range of these distributions according to *convergence* (60 for convergence, 600 for divergence). This makes “unlucky” draws in which conditions are visually not distinguishable less likely.

For each set, we also vary the value on the prediction scale (from 1000 to 2000) across which the range is centered. Considering our within-participant design, this makes it more likely that participants understand each set of predictions as being the result of a different stock, with a different true value. In order to assure all random draws from the distributions will appear on the response scale, we constrain the center of ranges of the uniform distributions to lie between 1300 and 1700. We define four centers–one per set of predictions–that divide this interval in (rounded) quartiles (1300, 1430, 1570, 1700). Given a center and a range, we then draw the predictions from uniform distributions (see next section). For example, in a draw for a divergent set of estimates with a center of 1700, each value within a range of 1400 and 2000 is equally likely to get selected. To avoid that single draws overlap too much within the same set, we defined a minimum space of 5 between the three predictions of a set.

For each set of predictions, we calculate the empirical mean based on the three drawn values. In the conflict of interest condition, this mean is the value that participants will be told experts gain when presented with the respective set of predictions. Consequently, the convergent predictions converge around what is said to be the “incentivized” value of the experts.

The same four sets of estimates will be used across the between-participant factor *independence*. The only thing that will differ between participants is the description of the context before seeing the sets of estimates.

Randomization. In a first series, we will pair the center values with ranges (as given by the *convergence*) such that each condition appears once in each half of the scale and each condition gets one of the extreme values (e.g. 1300 & convergence; 1430 & divergence; 1570 & convergence ; 1700 & divergence). We then randomly draw the four sets of predictions from their respective uniform distributions. In a second series, we proceed in the same way, but we inverse the pairs (i.e. 1300 & divergence; 1430 & convergence; 1570 & divergence ; 1700 & convergence). We randomly assign participants to one of these series.

Additionally, for each participant, we randomize the order of appearance of the sets of predictions within the respective series.

Dependent variables. For each set of estimates participants respond to the following questions: First, we ask participants about how accurate they deem the predictions ("On average, how accurate do you think these three predictions are" on a 7-point Likert scale ("not accurate at all" to "extremely accurate"). Second, participants are asked how they perceive the competence of the group of experts whose predictions they saw. Competence ("On average, how good do you think these three experts are at predicting the value of stocks?") is assessed on a 7-point Likert scale (from "not good at all" to "extremely good").

Statistical analysis

H1a and H1b: We will restrict the sample to those participants of the *independence* condition. We will then perform a paired t-test to assess the effect of *convergence* on *competence/accuracy*. This test accounts for the dependency between the measures across conditions issued by our within-participant design. [R code: *t.test(competence ~ convergence, data = independence_sample, paired = TRUE))*]

As a robustness check, we will run a mixed model OLS regression in which we regress *competence/accuracy* ratings on convergence. The model will include random intercepts for participants. [R code: *lmer(competence ~ convergence + (1 | participant), data = independence_sample)*]

H2: This hypothesis addresses an interaction: We want to know how the between factor *independence* alters the effect of the within factor *convergence*. We run an OLS regression with both convergence and independence and their interaction as independent variables, and competence (confidence) ratings as dependent variables. We have

competence = a + b1 convergence + b2 independence + b3 independence*convergence

where b1 is the effect of convergence, conditional on dependence, b2 is the effect of independence, conditional on divergence, and b3 the interaction term, indicating the difference of the effect of convergence between the independence and dependence condition. Based on these coefficients, we can calculate the expected means of the competence [confidence] ratings in the four conditions:

E(competence | convergence = 1 , independence = 0) = a + b1

E(competence | convergence = 0 , independence = 0) = a

E(competence | convergence = 1 , independence = 1) = a + b1 + b2 + b3

E(competence | convergence = 0 , independence = 1) = a + b2

If the interaction term b3 is significant, then there is a difference in the effect of convergence between independent and dependent contexts. If the term is positive, the effect is more positive in the independent context. [R code: *lm(competence ~ convergence + independence + convergence*independence, data = sample))*]

As a robustness check, we will run a mixed model OLS regression that will additionally include random intercepts and random slopes for participants. This is to control for the dependency of our observations, because each participant gives several convergence, but also several divergence ratings. Since we have one between (*independence*) and one within (*convergence*) factor, random slopes will only be included for the within factor, i.e. *convergence*. In this model, in order to assign meaningful random effects, we will change the *dummy coding* (0,1) for our factors to *effect coding* (-0.5, 0.5). With the effect coding, the interpretation of the regression coefficients changes, except for the interaction term:

> a = grand mean, i.e. the mean of competence ratings across all conditions
>
> b1 = main effect of convergence, i.e. average difference for within conditions convergence - divergence across levels of the factor *independence*
>
> b2 = main effect of independence, i.e. average difference for between conditions independence - conflict of interest across levels of the factor *convergence*
>
> b3 = interaction term, i.e. difference in effect of *convergence* between the two levels of *independence* (independence and conflict of interest)

[R code: *lmer(competence ~ convergence + independence + convergence*independence + (1 + convergence | participant), data = sample)*]

RQ1 and RQ2: We will restrict the sample to those participants of the *dependence* condition. We will then perform a paired t-test to assess the effect of *convergence* on *competence/accuracy*. This test accounts for the dependency between the measures across conditions issued by our within-participant design. [R code: *t.test(competence ~ convergence, data = dependence_sample, paired = TRUE))*]

As a robustness check, we will run a mixed model OLS regression in which we regress *competence/accuracy* ratings on convergence. The model will include random intercepts for participants. [R code: *lmer(competence ~ convergence + (1 | participant), data = dependence_sample)*]

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. Because we could build upon data from the first experiment, we ran a power analysis by simulation. The simulation code is available on OSF. For the simulation, we assumed the following means of competence ratings (on a scale from 1 to 7) in the four conditions of our study:

- mean_convergence_dependence = 4,
- mean_divergence_dependence = 4,
- mean_convergence_independence = 4.5,
- mean_divergence_independence = 3.5

The means for the last two lines correspond to the means of convergence and divergence, respectively, given independence. These were taken from experiment one. In fact, the mean for the convergence condition given independence in experiment one was even higher, namely 4.75. Since we want to be sure not to overestimate the effect of convergence in the power analysis, we chose 4.5.

We chose the means of the first two lines (convergence and divergence given dependence) to be the same because we didn't make any predictions about an effect of convergence given dependence. We chose the means to be 4 because that's the middle of the scale, representing a neutral response.

To model the participant dependent random effects in our data, we made the following assumptions (all standard deviations (SD) are to be interpreted relative to the 1-7 scale, they are not standardized) :

- by-subject random intercept SD = 0.6
- by-subject random slope SD = 1
- correlation between intercept and slope by-subject = 0.1

We further assumed the residual variation to, i.e. the variation in the data after accounting for all fixed and random effects, to be :

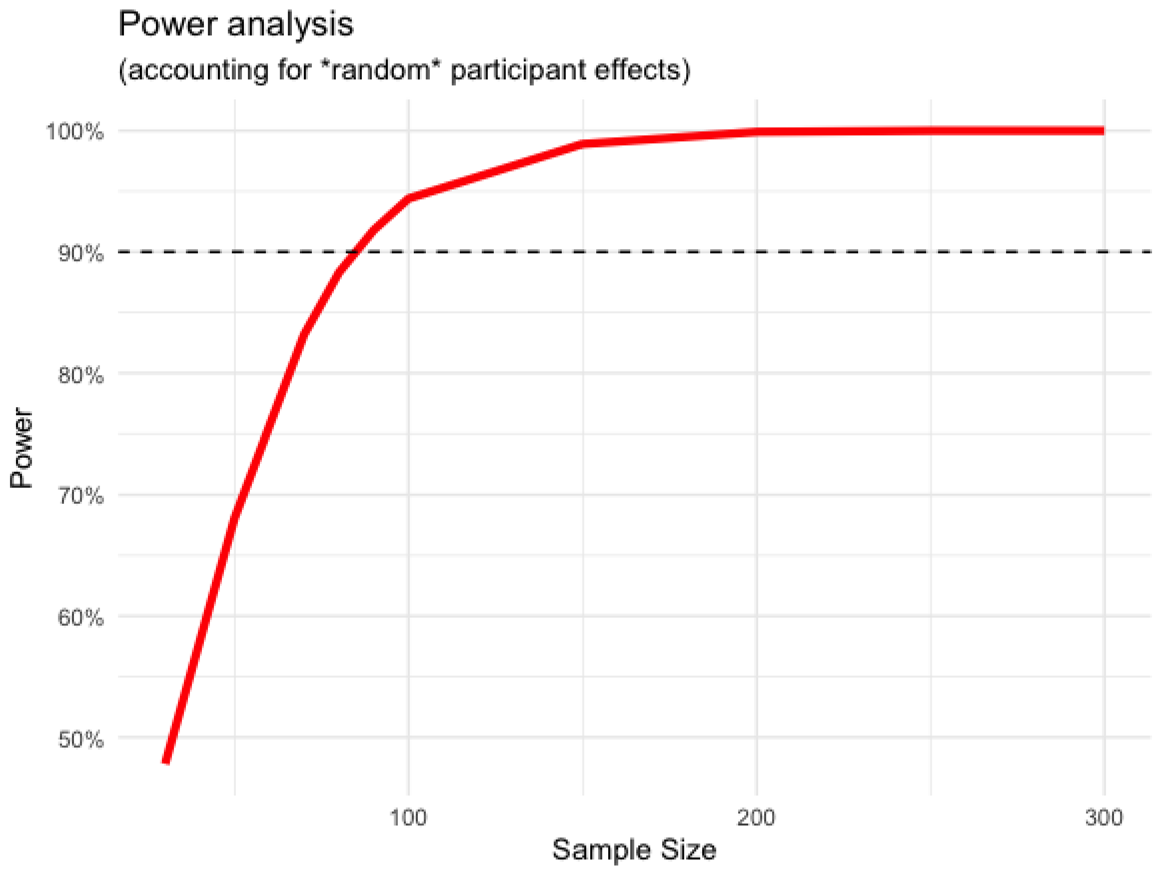
- residual SD = 1

We made all these assumptions based on data from the first experiment.

Choosing these means, we simulate competence ratings based on the following formula (written in R code):

```
competence = beta_0 + random_slope + (beta_1 + random_intercept) *
convergence_eff + beta_2*independence_eff + beta_3 * convergence_eff *
independence_eff + residual error
```

We set the power threshold high, at 90%. Simulating 1000 draws for various sample sizes between 20 and 300 {20, 30, 40, 50, 75, 100, 150, 200, 250, 300} we find that 100 participants provide a significant interaction term around 95% of the time (alpha for significance = 0.05).



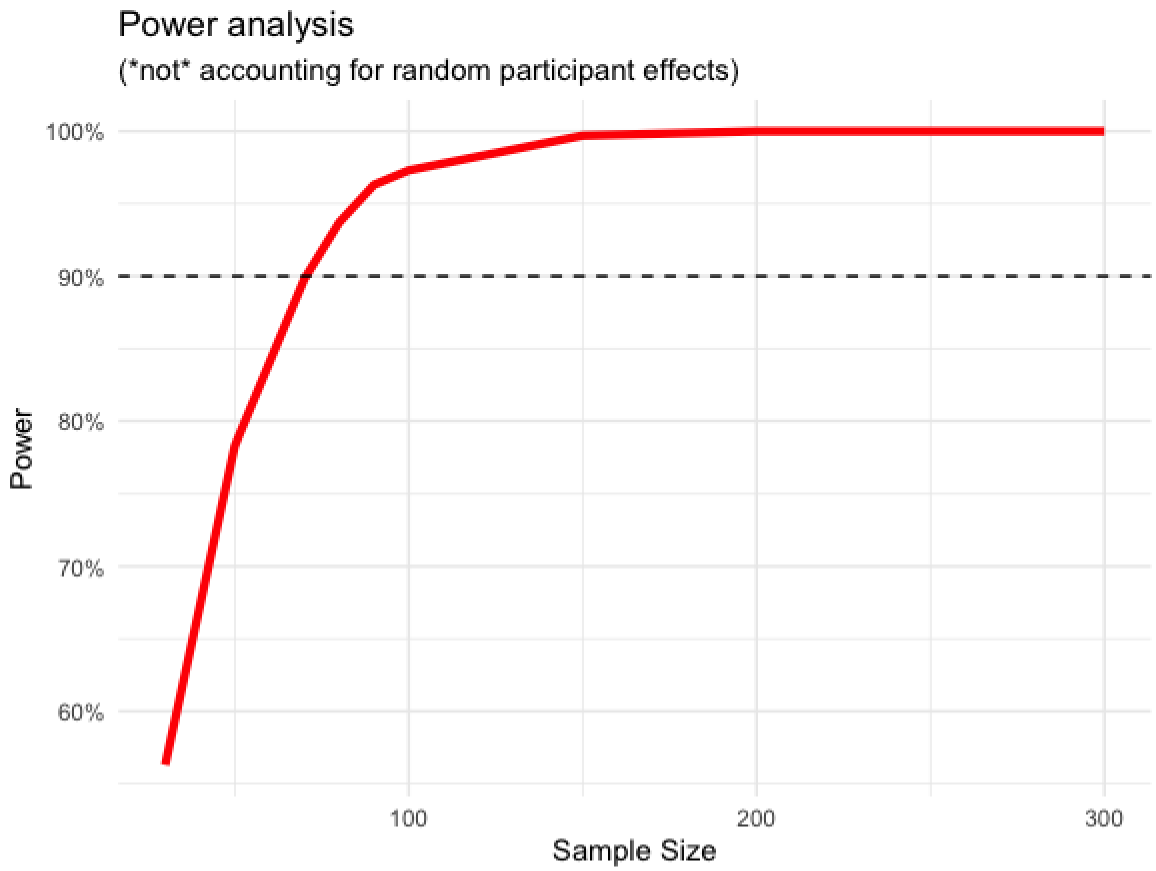
Due to uncertainty about whether some of the assumptions above will reflect the actual data generating process, we decided to run a seperate power simulation for reference. In this second simulation, we generated data without taking into account within-participant dependencies of data points (i.e. without including random effects). In other words, each observation (four per participant) was treated as an independent data point.

Choosing the same assumptions about means of the four conditions above, we simulate competence ratings based on the following formula (written in R code): :

```
competence = alpha + beta_1*convergence + beta_2*independence +
        beta_3*independence*convergence + rnorm(N, sd = residual_sd) ,
```

with the residual_sd (the residual standard error) being set to 1.3. This value is taken from a simple linear regression model from the first experiment, regressing competence on convergence.

We set the power threshold high, at 90%. Simulating 1000 draws for various sample sizes between 20 and 300 {20, 30, 40, 50, 75, 100, 150, 200, 250, 300} we find that 100 participants provide a significant interaction term around 97% of the time (alpha for significance = 0.05).

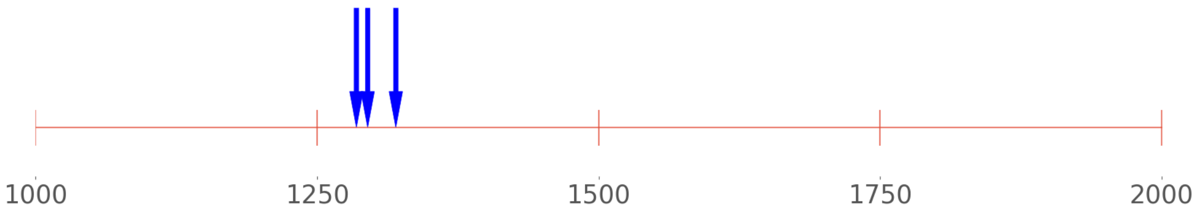


Both power analyses yield similar results. The assumptions that we based the simulations on were supported by data from experiment one; yet, the expected residual errors we used were very small compared to the expected effect size. Since we are not certain to observe such small (relative to the effects) residual errors in our interaction model in this experiment, we will increase the sample size compared to what the simulations suggest: we will double the number of participants and recruit 200 for this study.

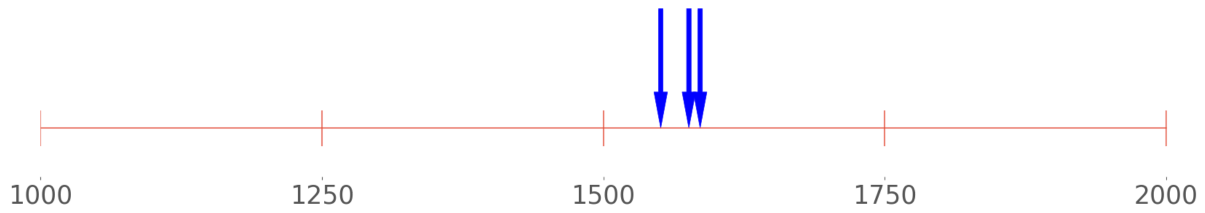
Appendix

Series I – sets of predictions:

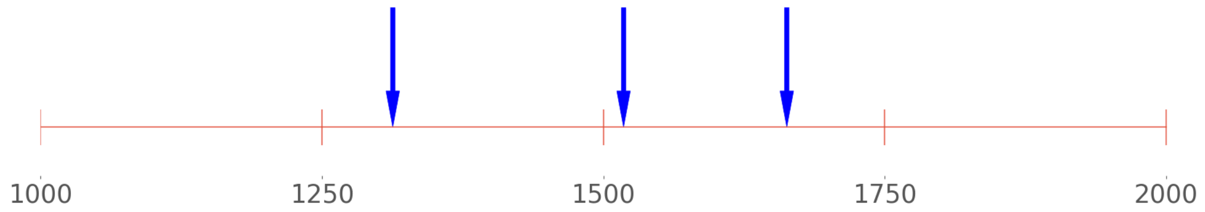
1. Convergence
    a. distribution center = 1300; empirical mean = 1300



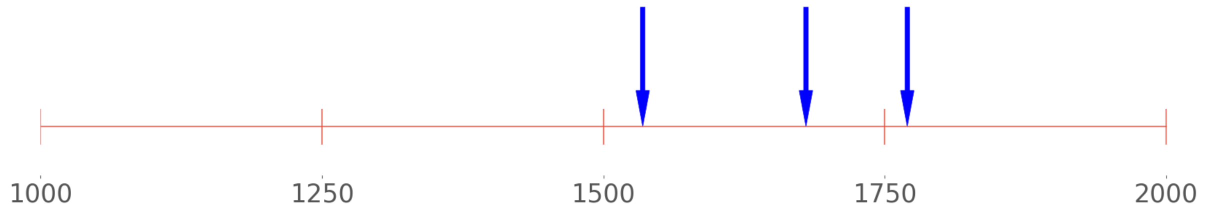
    b. distribution center = 1570; empirical mean = 1570



2. Divergence
    a. distribution center = 1430; empirical mean = 1500

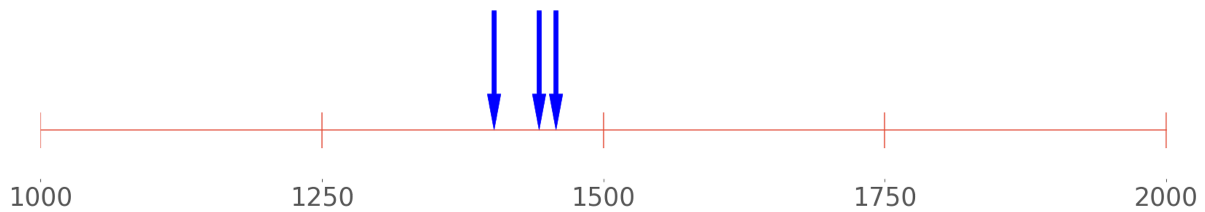


    b. distribution center = 1700; empirical mean = 1660

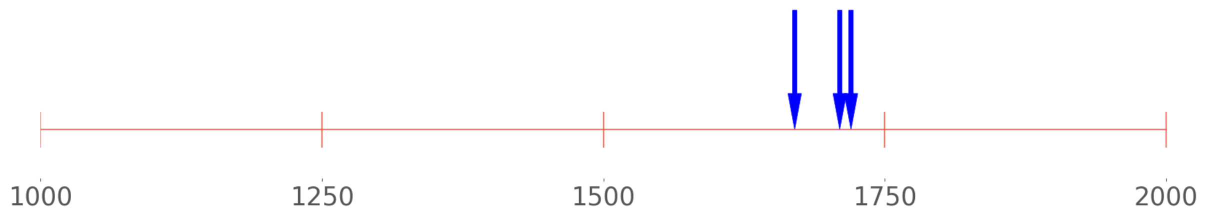


Series II – sets of predictions:

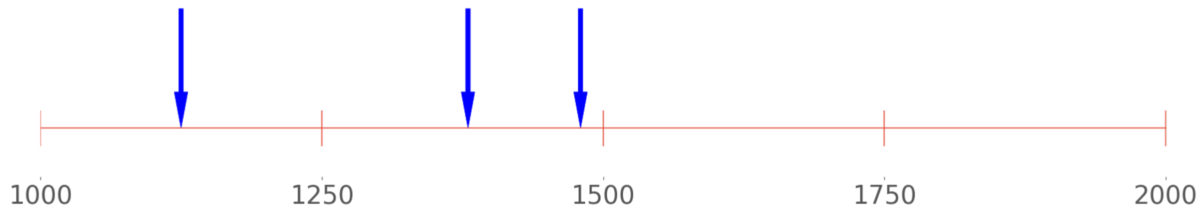
3. Convergence
    a. distribution center = 1430; empirical mean = 1430



    b. distribution center = 1700; empirical mean = 1700



4. Divergence
    a. distribution center = 1300; empirical mean = 1330



b. distribution center = 1570; empirical mean = 1580