Who Reviews The Reviewers? A Multi-Level Jury Problem

Ben Abramowitz
Tulane University
USA
babramow@tulane.edu

Omer Lev
Ben-Gurion University of the Negev
Israel
omerlev@bgu.ac.il

Nicholas Mattei
Tulane University
USA
nsmattei@tulane.edu

ABSTRACT

We consider the problem of determining a binary ground truth using advice from a group of independent reviewers (experts) who express their guess about a ground truth correctly with some independent probability (competence) p_i. In this setting, when all reviewers are competent with $p \geq 0.5$, the Condorcet Jury Theorem tells us that adding more reviewers increases the overall accuracy, and if all p_i's are known, then there exists an optimal weighting of the reviewers. However, in practical settings, reviewers may be noisy or incompetent, i.e., $p_i \leq 0.5$, and the number of experts may be small, so the asymptotic Condorcet Jury Theorem is not practically relevant. In such cases we explore appointing one or more chairs (judges) who determine the weight of each reviewer for aggregation, creating multiple levels. However, these chairs may be unable to correctly identify the competence of the reviewers they oversee, and therefore unable to compute the optimal weighting. We give conditions on when a set of chairs is able to weight the reviewers optimally, and depending on the competence distribution of the agents, give results about when it is better to have more chairs or more reviewers. Through simulations we show that in some cases it is better to have more chairs, but in many cases it is better to have more reviewers.

KEYWORDS

Peer Review, Peer Selection, Jury Theorem

ACM Reference Format:
Ben Abramowitz, Omer Lev, and Nicholas Mattei. 2025. Who Reviews The Reviewers? A Multi-Level Jury Problem. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 123 pages.

1 INTRODUCTION

People have been struggling with finding the *correct* answer for millennia[1]. In ancient times, when faced with a problem that required discovering a ground truth, two main approaches dominated. The first, less common today, was to approach deities and either ask them to intervene on the randomness of the world (as in the Book of Joshua, Chapter 7), which is a bit akin to sortition [18]; or to ask the deity's wisdom directly (e.g., the Oracle at Delphi). The second approach, still in widespread use today, is to try to assess the known information and draw a conclusion. This can either be done by laymen, the basic premise of the jury system as established by Magna Carta, or by people with expertise. In both cases, groups of people are used (instead of single individuals) to increase the reliability and accuracy of the answers, building on the "wisdom of the crowds".

Mathematical analysis of using a group of agents – a *jury*, or a set of *experts* – to assess information and make a decision has been done since at least Condorcet's time, in the late 18th century, when he established the Condorcet Jury Theorem (CJT) [15, 16]. In the standard jury setting, agents vote on a binary ground truth and the objective is to aggregate their votes, using a voting rule, to maximize the probability of the outcome being correct. It is typical to assume that agents guess the ground truth correctly with some independent probability (competence) p_i, we call agents competent when $p > 0.5$ and incompetent when $p \leq 0.5$[2]. According to the CJT, when the agents are competent, the collective accuracy of their majority vote tends to correctness as the number of agents increases. Even with a relatively small number of highly competent agents, accuracy can be very high. However, this result, which basically tells us that groups are less prone to mistake than individuals, rests on a knife's edge. By symmetry, if the agents are even minimally incompetent then, as the population grows, their collective accuracy under majority voting tends to 0, and a small group of highly incompetent agents stands no chance.

In the world around us, this idea is used everywhere – in judicial settings (juries), in academic conferences (peer evaluation), in voting for political leaders or in referendums, and even in settings with inanimate agents, such as aggregating sensor outputs into a single reading or indicator.

The precariousness of the CJT stems from the underlying aggregation procedure, majority voting, being anonymous, thus treating all agents equally, regardless of their competence. When agent competences can be different, majority rule is generally sub-optimal, and if one knows the agents' level of expertise exactly, the optimal aggregation method for maximizing accuracy with any number of independent experts and any competences is to use a weighted majority rule in which each agent's weight is the log-odds of their competence [28, 35]. In this result, somewhat surprisingly, the optimal weight of each agent does not depend on the competences of the other agents or even on the total number of agents. That is, if agents are added or removed, we do not need to update the weights in order for the weighted majority rule to remain optimal. However, the assumption that each agent's competence is known exactly is highly unrealistic.

In this work, we consider a variant of the classic jury setting, inspired by the domain of academic peer review [34], which attempts to address these unrealistic assumptions. Since the quality of reviewers may not be known by the conference Program Chairs, many conferences (e.g., IJCAI) appoint more senior researchers as SPCs (or chairs) to evaluate the reviewers and decide how to aggregate their views. This multi-level process inspired our model:

[1] Fans of *The Hitchhiker's Guide to the Galaxy* know it is **42**.

[2] The term competent is not meant to express a value judgment.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), A. El Fallah Seghrouchni, Y. Vorobeychik, S. Das, A. Nowe (eds.), May 19 – 23, 2025, Detroit, Michigan, USA.

we have both reviewers, who we will call *experts*, but also chairs, who we will call *judges*, that evaluate the experts and assign them weights.

Analyzing this setting is particularly interesting when the number of agents is relatively small, and therefore we cannot rely on the asymptotic guarantees of the CJT; as well as when there is a potential for significant deviation among agent competency, even when the particular competence values are unknown; or when agents can be incompetent, i.e., they will make the wrong decision most of the time[3]. We examine when such a two-level system works well, under what conditions it might be worthwhile to implement it, and when is it better to have an expert become a judge.

Contribution. We propose and investigate a novel model of multi-level jury problems for use when we have a small number of possibly unreliable agents. We show that when we know the agent competences exactly (or even approximately), we can find an optimal aggregation procedure, as long as the judges are competent. When the agents' (experts and judges) competences are unknown, we provide a set of numerical experiments demonstrating that adding more than a single judge is rarely helpful, and, indeed, in some cases, the potential damage of a less competent judge is enough to prefer to avoid judges completely.

2 RELATED WORK

There is a long history of studying the Condorcet jury model and its extensions (e.g., Ben-Yashar and Paroush [8], Berend and Paroush [10], Feld and Grofman [17], Grofman [20]), including work across computer science, philosophy, and economics. Our work is primarily based off the literature on weighting experts in both the offline [28, 35] and online settings [9, 14, 19, 41], though in this work we restrict our focus to a single decision. The overall CJT model can be seen in portfolio solver techniques where slower, more reliable algorithms evaluate ensembles of faster, less accurate algorithms [40] and in boosting techniques in ML, where one aggregates weakly competent classifiers into a better overall classifier [33]. These methods are particularly useful when each expert will use limited effort or energy [39].

In settings with repeated decisions, the competences of the experts can be estimated based on their performance history. Their competence might be estimated by their similarity to other agents, or how often they agreed with the decision outcomes in the past [4, 23, 32]. In our setting, however, we do not have access to this history and cannot use it to estimate competences. Indeed, in peer review, one may have a notion of other reviewers' competency, but rarely does one co-review with another to form a precise estimate.

Our emphasis on imperfect judges is also inspired by work on "wisdom of the crowds" and crowdsourcing [12, 13, 36], proxy voting [1, 30] and truth-tracking in Liquid Democracy [7, 44]. For example, in proxy/delegative voting the voters are like judges assigning weights (voting units) to their proxies/delegates. But, as noted above, a major inspiration has been the work on academic peer evaluation [34], in which experts assess each other's competences. Some treat this matrix as a Markov chain, and use its eigenvector values as the experts' weights [21] in a manner reminiscent of some peer-evaluation models [26, 29, 42]. In contrast, within our setting the agents who vote and the of agents who weight the voters are disjoint.

The problem of partitioning agents into judges and experts is also related to the problem of computing optimal committee sizes [27, 31]. There is also work on how group accuracy depends on the size of the group and mean competence [20, 22], which is reflected in our simulations.

3 MODEL AND NOTATION

Our model has two types of agents; judges and experts. Let E be a set of m experts and J be a set of n judges. The experts vote on a single binary issue where there is only one correct (ground truth) outcome. Without loss of generality, let the options be $\{1, 0\}$ where 1 is correct and 0 is incorrect. Each expert $e \in E$ has a *competence*, or probability p_e of voting correctly, independent of all other experts. We associate each expert's index with their vote, so expert $e \in E$ casts a vote $v_e \in \{1, 0\}$ with competence $p_e = P(v_e = 1)$. If an agent's competence is above 1/2 we will say that they are competent, and call them incompetent otherwise. We assume no one is always correct or always incorrect, and so $0 < p_e < 1$.

Weighted Majority Rules. We refer to the probability of producing the correct outcome as the *accuracy*, and reserve the term *competence* to refer to individual agents' probabilities of voting correctly; i.e. accuracy is collective.

Definition 1 (Weighted Majority Rule). *A weighted majority rule gives each expert $e \in E$ a weight $w_e \in \mathbb{R}$ and selects 1 as the winner if $\sum_{v_e=1} w_e > \sum_{v_e=0} w_e$, selects 0 as the winner if $\sum_{v_e=1} w_e < \sum_{v_e=0} w_e$, and uses a tie-breaking rule (e.g. coin flip) for the edge case where these sums are equal*[4].

Definition 2 (Simple Majority Rule). *Simple majority rule refers to the weighted majority rule where all weights are equal and positive and ties are broken randomly.*

The Condorcet Jury Theorem tells us that if $p_e \geq 0.5 + \epsilon$ for some $\epsilon > 0$ for all experts, then with simple majority accuracy tends to 1 asymptotically as the number of experts tends to infinity. A weighting function maps vectors of values in $(0, 1)$ (i.e. competences) to equal length vectors of real values. For any set of experts, including incompetent ones, the optimal aggregation method of experts' votes, is to apply the log-odds weighting function to the experts' competences and use the corresponding weighted majority rule [28, 35].

Definition 3 (Log-Odds Weighting Function). *Given a vector of values in the open unit interval $\vec{p} = (p_1, \ldots, p_m)$, the log-odds weighting function returns the vector $\vec{w} = (w_1, \ldots, w_m)$ where $w_e = \log(\frac{p_e}{1-p_e})$ for all $1 \leq e \leq m$.*

Any weighting of the agents implies a collection of winning coalitions – subsets of agents who, if they all vote the same way,

[3] In academic peer evaluation it is uncommon for reviewers to be given negative weights, which we will see is required for the log-odds rule to be optimal. But it is common in other settings, e.g., sensors or proxy voting scenarios where one might want to always do the opposite of a political rival.

[4] Note that ties never occurred in our simulations as the need for tie-breaking, with real valued weights, is highly unlikely.

determine the outcome regardless of the other votes [38]. Different weightings may yield the same rule because they imply the exact same winning coalitions. For example, with 5 agents there are exactly 7 distinct weighted majority rules [24, 25]. Multiplying the weights of all agents by a constant does not change the winning coalitions and therefore does not change the rule. Similarly, perturbing agent weights by small amounts may not change the winning coalitions. Therefore, while the weights may vary continuously, the accuracy under various weightings will change in discrete steps. In practice, weights may be finite precision rather than true real numbers, and this is also the case in our simulations that use floating point arithmetic, but as long as the rounding tends to be too small to change winning coalitions for most instances its effect will be negligible.

The log-odds weighting rule assigns a positive weight to competent experts when $p_e > 0.5$, weight of zero if $p_e = 0.5$, and negative weight to any incompetent expert with $p_e < 0.5$. In some settings it may be inappropriate to allow negative weights and better to assume any such weights are rounded up to zero. Bounding weights below by zero has the effect of ignoring the incompetent experts and is therefore qualitatively similar to assuming all experts are competent, though with a smaller number of experts. We therefore focus on the more informative setting where weights can be negative. Negative weights also have real-world motivation. A remote sensor may have drifted so far off its initial calibration to be *reliably* wrong, as has happened with many spacecraft [6]. However, we would like to believe that peer reviewers, jurors, and the like are not so reliably wrong that negative weights would be needed.

Multi-Level Jury Problems. Each judge $j \in J$ estimates the competence of each expert $e \in E$ as $p_{je} \in (0, 1)$ and assigns them a weight w_{je} based on this estimate. When assigning weights to the experts, our judges always use the log-odds weighting function on their competence estimates. Intuitively, our judges are trying to implement the optimal rule using their estimates of the experts' competences. Formally, judge j assigns expert e a weight $w_{je} = \log(\frac{p_{je}}{1-p_{je}})$. When there are multiple judges, the weight of an expert will be the average weight assigned to them by the judges $w_e = \frac{1}{n}\sum_{j} w_{je}$.

Perceived Competences. Our theoretical results depend on judges using the log-odds weighting but do not depend on how the judges *form* their competence estimates p_{je}. To perform empirical analysis we must make assumptions about where these estimates come from. Rather than drawing the estimates p_{je} from some named distribution with mean p_e, we take an approach inspired by peer review, and assume that judges and experts are fundamentally similar. Each judge j has competence p_j just like the experts and estimates the competence of expert e as $p_{je} = (p_j \cdot p_e) + (1 - p_j)(1 - p_e)$, i.e., the probability that expert e agrees with them. When p_{je} is derived in this way, we will refer to it as the judge's *perceived competence* of the expert. As with peer review, a judge may estimate an expert's competency from knowing them professionally, but may not have observed many past reviews. A judge could also reach this estimate of competency if they observe enough votes from the expert but the ground truth is never revealed, as is the case in some peer prediction settings [43]. While there are many models one could select for the formulation of perceived competencies, this formulation does not rely on complex models of estimation or opinion formation. Indeed, any additional assumptions that improves accuracy, i.e., that judges are better at estimation than simple observation, would improve the limit results we obtain. So while this assumption is simplistic, it is in some ways a worst-case assumption for our theoretical results, and leads to interesting empirical results.

Example 1. *Suppose we have 5 experts with competences* $\vec{p}_E = (0.6, 0.6, 0.6, 0.7, 0.9)$. *The optimal log-odds weighting is approximately* $\vec{w}^*_E = (0.41, 0.41, 0.41, 0.85, 2.2)$. *With these weights the most competent expert* ($p_e = 0.9$) *receives a weight* ($w_e = 2.2$) *that makes them a dictator in a weighted majority vote, since their weight is greater than all other experts combined. Hence, the accuracy under the log-odds weighting is exactly 0.9. If instead we use simple majority, the accuracy drops to 0.82.*

A judge with $p_j = 0.6$ *would assign the experts weights of approximately* $\vec{w}^{0.6}_E = (0.08, 0.08, 0.08, 0.16, 0.323)$ *using the log-odds weighting of perceived competences. Note that the fifth expert is no longer a dictator. How high of a competence would the judge need to have to assign weights that results in the optimal weighting? The judge's competence would have to be greater than* 0.962*; a far cry from* 0.6 *and higher than all the experts. And yet, the judge's sub-optimal weighting yields an accuracy of 0.898, which is a great improvement over simple majority, and extremely close to optimal at 0.9! Example 1 is illustrated in Figure 1 where we plot the accuracy, sweeping* p_j *from 0.0 to 1.0.*

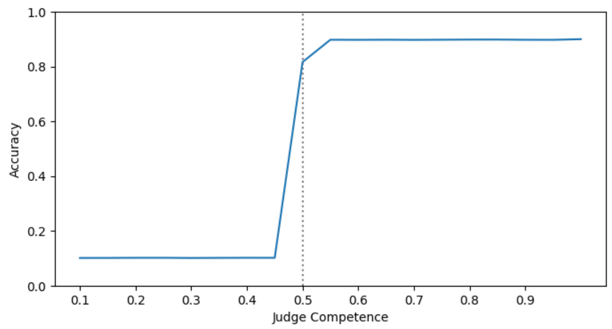


Figure 1: Accuracy of perceived optimal weightings from a single judge with the expert competences as in Example 1.

In Example 1 we rounded the values of the weights to two decimal places, which did not change the rule implemented. Similarly, whether judges are human, sensors, or algorithms, they do not (always) need to provide high precision of weights. More specifically, the smaller the number of agents, the less chance there is for small perturbations or rounding of the weights to change the rule. This is also why Figure 1 will be piecewise linear regardless of the step size.

4 OPTIMALITY AND ROBUSTNESS

With a single judge, if $p_{je} = p_e$ for all $e \in E$, then all experts receive their optimal weight. As noted, small perturbations to the weights do not change the rule because the winning coalitions determined by the weights do not change. Thus, if p_{je} is close enough to p_e for all experts, they will still produce the optimal weighting. We now establish sufficient conditions for an ensemble of judges to produce the optimal weighting, and provide a condition under which the difference between an expert's weight and their optimal weight tends to be small.

Proposition 1 shows that when the geometric mean of the judges' perceived competence of experts odds is their true competence odds, all experts are assigned their optimal weights, $w_e = w_e^*$. This does not require individual judges to know the experts' true competences, and does not depend on the number of experts nor the number of judges.

Proposition 1. *If each judge uses the log-odds weighting function on their estimates of expert competences, and the geometric mean of the judges' estimates of each expert's competence odds is the expert's true odds, then the weighted majority rule using the judges' average weights to weight each expert is exactly the optimal weighted majority rule.*

Proof. Since judge j gives each expert a weight of $w_{je} = \log(\frac{p_{je}}{1-p_{je}})$,

$$w_e = \frac{1}{n}\sum_j w_{je} = \frac{1}{n}\sum_j \log(\frac{p_{je}}{1-p_{je}}) =$$
$$= \frac{1}{n}\log(\prod_j \frac{p_{je}}{1-p_{je}}) = \log((\prod_j \frac{p_{je}}{1-p_{je}})^{\frac{1}{n}})$$

We assume the geometric mean of judge' estimates of the experts' competence odds is correct, i.e., $(\frac{p_e}{1-p_e}) = (\prod_j \frac{p_{je}}{1-p_{je}})^{\frac{1}{n}}$. Thus, $w_e = \log(\frac{p_e}{1-p_e}) = w_e^*$. □

This result requires assumptions on the judges' competence estimates that must hold for *all* experts. However, suppose there are errors in these collective competence estimates. We want to know how sensitive the weight of a single expert is to such errors.

Corollary 1. *If the geometric mean of judge estimates of competence is off by some multiplicative factor α for some expert, then the error of that expert's weight is only* $\log(\alpha)$.

Proof. In the proof above, assume instead that $\alpha\left(\frac{p_e}{1-p_e}\right) = (\prod_j \frac{p_{je}}{1-p_{je}})^{\frac{1}{n}}$. Then $w_e = \log(\alpha \cdot \frac{p_e}{1-p_e}) = w_e^* + \log(\alpha)$. □

Admittedly, it is not clear in what settings the conditions of Proposition 1 should be expected to hold. Neither can we make claims about what multiplicative factors are realistic in Corollary 1. But ultimately, what we care about most is the sensitivity of the accuracy to errors in competence estimates, which has to do with the set of winning coalitions induced by the weights. We care less about the sensitivity of the weights themselves, although the sensitivity of the weights gives some intuition.

In Example 1, we see that for all $p_j > 0.55$ the accuracy rivals that of the optimal rule, with almost no difference. The effect that dominates Figure 1 is when we move from $p_j < 0.5$ to $p_j = 0.5$ (when the rule becomes simple majority), with another slight bump with a move to $p_j > 0.55$.

Recent work [5] shows that when expert competences are drawn from certain distributions over the range $(1/2, 1)$, simple majority achieves an accuracy close to optimal. However, as one might expect, when experts can be incompetent the majority rule is no longer a good approximation to the optimal weighted majority rule. Thus, if judges can at least differentiate the competent from incompetent experts, the weighting they produce should be expected to outperform simple majority rule when there are incompetent agents. In our model, any minimally competent judge with $p_j > 0.5$ is able to achieve this. We will discuss this more in Section 5 with an array of experiments, including a suite of experiments with a single judge. For now we introduce some basic theoretical observations that help understand the phenomena observed in our experiments. Namely, the improvement in accuracy as the judge(s) competence(s) reach $p_j = 0.5$ appears to be largely but not entirely explained by the judge's ability to (1) assign experts weights of the correct sign, and (2) assign experts weights according to the weak order of their competences.

Example 2 (Two Experts). *Suppose there are two experts with competences (p_1, p_2) such that $p_1 > p_2$. If $p_2 > 1/2$, i.e., both experts are competent, the optimal aggregation rule is to make p_1 dictator. However, if $p_1 > 1/2 > p_2$ then the optimal rule is either to make p_1 a dictator if $p_1 \geq 1 - p_2$, or else make p_2 an anti-dictator using negative weight such that the outcome is the opposite of however p_2 votes. If $1/2 > p_1 > p_2$, then the optimal rule makes p_2 the anti-dictator symmetrically with the first case.*

From Example 2, we see that even with only two experts, if a judge can determine which experts are competent, and order their competences correctly, this is enough information to produce the optimal rule. With more experts, the situation is more complicated, but our experiments reveal that merely separating the competent experts from incompetent ones creates a large improvement in overall accuracy. Any chair with $p_j > 1/2$ using log-odds weightings of the perceived competences can achieve this improvement.

Proposition 2 (Correct Sign). *If $sign(p_{je} - 0.5) = sign(p_e - 0.5)$, then $sign(w_{je}) = sign(w_e^*)$.*

Proof. Proposition 2 follows directly from the fact that $\frac{p}{1-p} > 1$ if and only if $p > 1/2$, and therefore $\log(\frac{p}{1-p}) > 0$ if and only if $p > 1/2$. Symmetrically for $p < 1/2$. □

Proposition 3 (Correct Order). *If p_{je} is a strictly monotonic increasing function of p_e, then the weak order of expert weights given by judge j is the weak order of the experts' competences.*

Proposition 3 follows from the monotonicity of the log-odds weighting. When a single judge applies the log-odds weighting to their perceived competences of a small set of experts then we can make the following observations.

Observation 1. *If $p_j = 1/2$ then all experts are equally weighed. If $p_j = 1$ then experts are optimally weighed.*

If $p_j = 1/2$ then the judge will perceive all experts as having a competence of $1/2$, and therefore assign them the weight of 0, which we treat as giving them equal weight, i.e. majority vote. When $p_j = 1$, the judge knows exact competences of the experts and therefore assign them their optimal weights.

Observation 2. *If $p_j > 1/2$, then $p_{je} > p_{je'}$ iff $p_e > p_{e'}$.*

This means that a judge's perceived competences of the experts preserves the order of their true competences if $p_j > 1/2$. When weights are based on perceived competences, this means whenever $p_j > 0.5$, the judge will assign all experts' weights with the correct sign and in the correct order. This is because when $p_j > 0.5$, p_{je} is monotonically increasing in p_e and $p_{je} > 0.5$ iff $p_e > 0.5$. Thus, even a single barely competent judge might give us an edge over simple majority.

Theorem 1 (Minimal Competent Single Judge). *If $p_j > 0.5$ and the judge assigns experts weights according to their perceived competences, the weights given by the judge will have the correct sign and the correct order.*

Proof. Let $p_j = 1/2 + \varepsilon_j$ and $p_e = 1/2 + \varepsilon_e$.

$$\begin{aligned} p_{je} &= \left(1/2 + \varepsilon_j\right)\left(1/2 + \varepsilon_e\right) + \left(1/2 - \varepsilon_j\right)\left(1/2 - \varepsilon_e\right) \\ &= 1/2 + 2\varepsilon_j \varepsilon_e \end{aligned}$$

If $\varepsilon_j > 0$ and $\varepsilon_e > 0$, then this value is greater than $1/2$, if $\varepsilon_j > 0$ and $\varepsilon_e < 0$, then this value is less than $1/2$, and if $\varepsilon_e = 0$ then this value is exactly $1/2$. The proposition then follows from Proposition 3 and Proposition 2. □

We can generalize Proposition 2 by replacing the requirement that $p_j > 1/2$ with the requirement that the geometric mean of judges' estimated competence odds is greater than 1. Notice that the requirements for this theorem are far weaker than for the optimality demanded by Proposition 1.

Proposition 4 (Correct Sign). *If the geometric mean of every expert's estimated competence from the judges is greater than* 1 *whenever $p_j > 1/2$, less than* 1 *whenever $p_j < 1/2$, equal to* 1 *when $p_j = 1/2$, every expert will be assigned a weight with the same sign as their optimal weight.*

Proof. Suppose $p_e > 1/2$. Their optimal weight is positive, and so we need the following to hold:

$$\sum_{j \in J} \log\left(\frac{p_{je}}{1 - p_{je}}\right) > 0$$

$$\prod_{j \in J} \frac{p_{je}}{1 - p_{je}} > 1$$

$$\left(\prod_{j} \frac{p_{je}}{1 - p_{je}}\right)^{\gamma} > 1$$

for $\gamma > 0$. When $\gamma = \frac{1}{n}$ this is the geometric mean. The case for $p_e < 1/2$ is symmetric with flipped inequality, and for $p_e = 1/2$ is the same but with strict equality. □

Proposition 2, Proposition 3, and Theorem 1 can be immediately generalized to multiple judges by requiring their respective assumptions to hold for all judges individually, because the sum of non-negative (strictly monotonic) functions remains non-negative (strictly monotonic). While our results on correct sign and order are interesting, they are not sufficient to always outperform equal weighing. As we will see in the next section, with our perceived competences model we see a discrete interpolation between equal and optimal weighting as a single judge's competence increases. Sign and order preservation appear to be part of the explanation as to why, but this explanation is incomplete. We pose the conjecture of monotonicity w.r.t. single judge competency using perceived competences as an open problem.

5 EMPIRICAL RESULTS

We gain a deeper understanding of the behavior we see in Example 1 and Section 4 with a set of numerical experiments. In our simulations, the agents' competences are drawn i.i.d from various distributions. All experiments were run for 100,000 iterations for each parameterization of the problem instance so that the variances are negligible. When there is no judge ($n = 0$), unweighted majority vote is used with random tie-breaking.

We consider three distributions of expert competences: uniform, truncated normal, and truncated exponential. The uniform distribution reflects settings where the experts can equally have any competence; the exponential distribution models settings where the expertise tends to be rare [11]; and the normal distribution is appropriate for a common expertise [37].

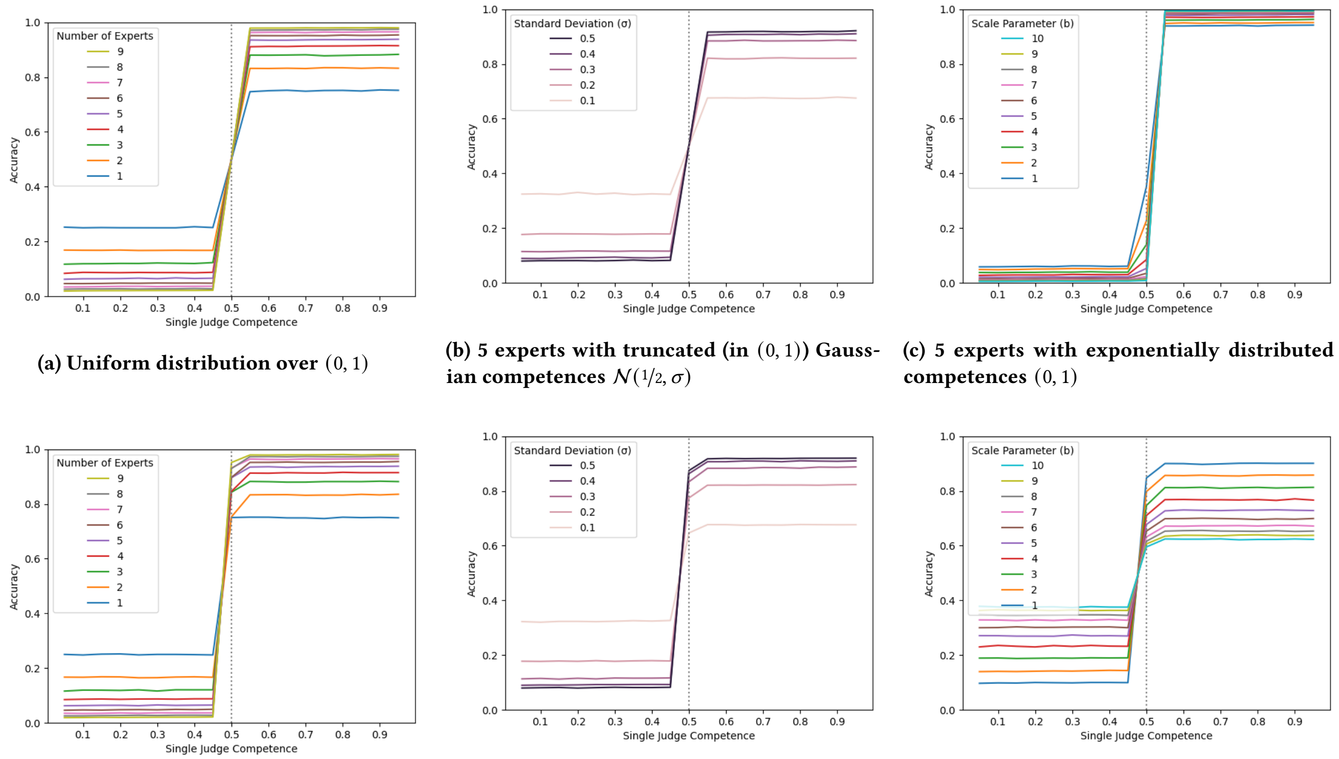
We provide Tables 1-3, with one for each family of competence distribution over $(0.5, 1.0)$ where no agents are incompetent and all agents, judges and experts, have their competences drawn from the same distribution. By looking across the rows, we see the improvement from incrementally adding experts, by looking down each column we see any change from adding judges, and by looking at the diagonals we see how best to choose n and m to divide the agents into roles given a fixed number of agents $n + m$.

When competences are bounded below by 0.5, we see that there is always a benefit to increasing the number of experts, but no benefit to increasing the number of judges beyond $n = 1$. Notably, without any judges, adding an expert only increases accuracy when incrementing the number of experts from even to odd. When adding a single expert makes m even, there is no observable increase in accuracy if there is no judge. With one or more competent judges, increasing the number of experts always increases accuracy. By contrast, adding a single judge shows increasing accuracy whenever there is more than 1 expert, but increasing the number of judges further shows no benefit. See Appendix for a suite of similar experiments varying the support and parameters of the various distributions.

In all three of these tables, there is a noticeable benefit to adding a single judge, but no benefit to adding judges beyond that. We therefore take a closer look in Section 5.1 at the case with a single judge, and examine how the accuracy varies if the judge's competence differs from the experts, even allowing the judge to be incompetent. Lastly, we will look in Section 5.2 at how to determine the number of agents to set as judges versus experts when all agent competences are drawn from the same distribution and agents can be incompetent. Further tables can be found in the Appendix.

5.1 Single Judge

The top row of Figure 2 shows accuracy as a function of the single judge's competence when expert competences are distributed over the interval $[0.001, 0.999]$ according to the uniform, truncated normal ($\mathcal{N}(1/2, \text{varying } \sigma)$), and truncated exponential distributions (using the density function $e^{-x}/1-e^{-b}$ for varying values of b) respectively. Only Figures 2a and 2b exhibit true symmetry because



(a) Uniform distribution over $(0, 1)$

(b) 5 experts with truncated (in $(0, 1)$**) Gaussian competences** $\mathcal{N}(1/2, \sigma)$

(c) 5 experts with exponentially distributed competences $(0, 1)$

(d) Uniform distribution over $(0.5, 1)$

(e) 5 experts with truncated (in $(0.5, 1)$**) Gaussian competences** $\mathcal{N}(1/2, \sigma)$

(f) 5 experts with exponentially distributed competences in $(0.5, 1)$

Figure 2: Accuracy with a single judge and expert competences drawn i.i.d. from a distribution with support $[0.001, 0.999]$ **(top row) or support** $[0.501, 0.999]$ **(bottom row).**

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.750	0.749	0.843	0.843	0.897	0.897	0.929	0.929	0.950	0.950
1	0.750	0.831	0.880	0.913	0.936	0.951	0.964	0.973	0.980	0.984
2	0.751	0.833	0.881	0.914	0.937	0.953	0.964	0.974	0.980	0.984
3	0.748	0.833	0.880	0.914	0.937	0.952	0.964	0.972	0.980	0.984
4	0.748	0.832	0.882	0.914	0.937	0.952	0.964	0.974	0.980	0.984
5	0.747	0.833	0.882	0.915	0.937	0.953	0.965	0.972	0.980	0.984
6	0.748	0.832	0.881	0.914	0.937	0.952	0.964	0.974	0.980	0.985
7	0.750	0.833	0.881	0.913	0.937	0.954	0.964	0.974	0.980	0.984
8	0.748	0.834	0.881	0.915	0.937	0.952	0.964	0.974	0.980	0.985
9	0.751	0.832	0.881	0.914	0.937	0.953	0.965	0.974	0.980	0.984
10	0.748	0.833	0.881	0.915	0.937	0.954	0.966	0.974	0.979	0.984

Table 1: Accuracy with all agent competences drawn from the uniform distribution over (0.501,0.999)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.581	0.578	0.621	0.618	0.647	0.648	0.673	0.669	0.693	0.692
1	0.580	0.616	0.639	0.658	0.673	0.691	0.705	0.720	0.732	0.741
2	0.580	0.612	0.640	0.660	0.680	0.690	0.705	0.719	0.729	0.743
3	0.581	0.612	0.640	0.659	0.676	0.691	0.707	0.718	0.730	0.742
4	0.581	0.614	0.638	0.658	0.673	0.692	0.707	0.719	0.727	0.741
5	0.581	0.616	0.635	0.658	0.674	0.692	0.706	0.718	0.729	0.743
6	0.580	0.612	0.635	0.660	0.675	0.692	0.706	0.717	0.731	0.742
7	0.579	0.613	0.639	0.659	0.678	0.692	0.707	0.718	0.730	0.741
8	0.581	0.615	0.639	0.659	0.676	0.689	0.708	0.718	0.732	0.741
9	0.582	0.613	0.639	0.659	0.675	0.692	0.707	0.719	0.733	0.741
10	0.579	0.612	0.641	0.662	0.676	0.692	0.705	0.718	0.731	0.743

Table 2: Accuracy with all agent competences drawn from the Gaussian distribution $\mathcal{N}$**(0.5,0.1) truncated over (0.501, 0.999)**

competences are drawn from a symmetric distribution with mean 0.5, and Figures 2d and 2e show behavior most similar to Figure 1.

In the top row of Figure 2 we can have highly incompetent experts, but even in this setting whenever the judge has competence $p_j > 1/2$, high overall accuracy is achieved. This is because the ability of the judges to differentiate competent experts from incompetent ones is of primary importance, and Proposition 4 shows that a judge using perceived competences is able to do this. In Figures 2a-2c, once the judge passes a minimum threshold of competence, little

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.672	0.672	0.750	0.747	0.797	0.799	0.834	0.835	0.862	0.865
1	0.674	0.745	0.795	0.831	0.857	0.881	0.897	0.913	0.925	0.935
2	0.672	0.744	0.795	0.829	0.858	0.882	0.898	0.912	0.926	0.936
3	0.675	0.744	0.796	0.829	0.858	0.879	0.899	0.913	0.925	0.936
4	0.671	0.747	0.791	0.830	0.858	0.880	0.897	0.913	0.924	0.938
5	0.672	0.746	0.794	0.831	0.857	0.879	0.897	0.912	0.925	0.936
6	0.672	0.746	0.791	0.829	0.858	0.880	0.898	0.913	0.926	0.935
7	0.674	0.742	0.793	0.831	0.857	0.880	0.896	0.913	0.925	0.936
8	0.672	0.745	0.792	0.829	0.858	0.880	0.897	0.913	0.925	0.937
9	0.671	0.747	0.796	0.829	0.858	0.881	0.898	0.913	0.925	0.936
10	0.674	0.745	0.795	0.832	0.857	0.878	0.897	0.912	0.927	0.936

Table 3: Accuracy with all agent competences drawn from the exponential distribution with scale parameter $b = 2$ truncated over (0.501, 0.999)

is gained from increasing p_j. Interestingly, in Figure 2b we see that when expert competences are distributed normally with mean 1/2, higher variance leads to higher collective accuracy. This appears to be because a judge with sufficiently high competence can differentiate between highly competent and minimally competent experts, and then leverage the benefits of having highly competent experts when they are present.

In the bottom row of Figure 2 we show accuracy as a function of the single judge's competence when expert competences are distributed over the interval $[0.501, 0.999]$ according to our three distributions. This is closer to prior work in the literature where all experts are assumed to be competent. Unlike in Figure 2a and 2b, which are symmetrical distributions, we now see an asymmetry around $p_j = 0.5$. When the judge's competence is 1/2, the judge gives all experts the same weight, so when experts competence is symmetrical around 1/2 (as in the upper row of Figure 2), resulting in an accuracy of 1/2, but when they cannot be incompetent, the resulting accuracy is higher – almost optimal [5]. In contrast to the top row of Figure 2, when all experts are competent, there is a large difference in accuracy for truncated exponential distributions with different scale parameters.

5.2 Should We Add a Judge or an Expert?

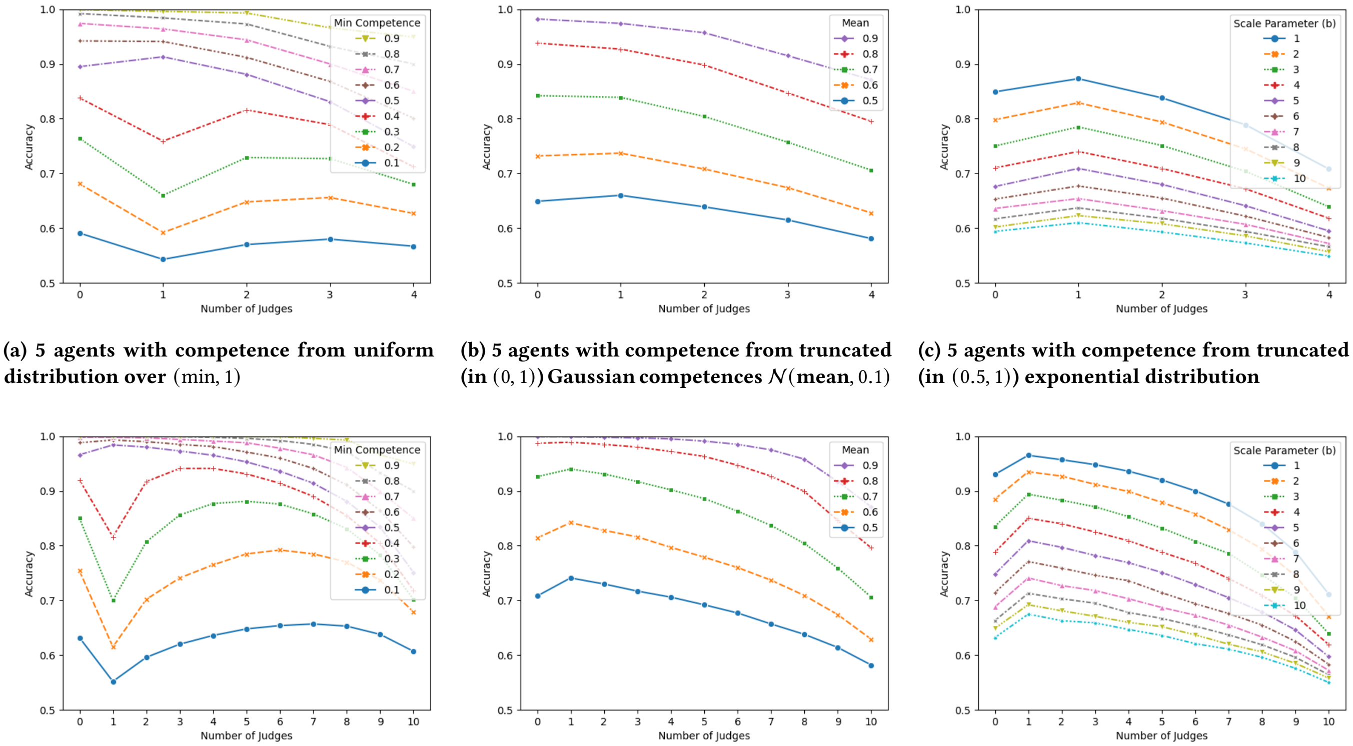
Empirically, with a single judge the accuracy improves as the judge's competence grows, and we know from the Condorcet Jury Theorems that as the number of experts increases, if the experts are competent, accuracy will increase. Hence, if the conditions of Proposition 4 hold, then accuracy will increase as the number of experts increases whether they are competent or incompetent, as long as they have competences that are not equal to 1/2. We examine the balance between the benefits of increasing the number of judges and increasing the number of experts. That is, with a fixed set of agents of unknown competences, how should they be partitioned between experts and judges? This problem is faced by any scientific conference with a hierarchical structure: how to divide its Program Committee between reviewers and SPCs, ACs, etc.

We first draw agents' competences from uniform distributions with varying lower bounds and examine the optimal number of agents to set aside as judges rather than experts when we have 5 and 11 agents, respectively (Figures 3a and 3d). For both number of agents, we see that *setting aside a single agent as judge diminishes the accuracy compared to the simple majority rule in almost all cases.* This is more pronounced when there is a possibility the judge will have competence below 1/2, i.e., when a lone judge is incompetent they give all competent experts negative weights and incompetent experts positive weights. The only case where a judge is helpful is when the minimum competence of agents is 1/2, perhaps because there is high enough chance that the judge will be helpful, and the agents' competence is not guaranteed to be high enough that losing the judge as an expert is too big a hit. Even adding more judges, at best, returns the accuracy to the level of a simple majority rule, though most commonly it does not. In the 11 agent case, Figure 3d, this effect is even more pronounced than for 5 agents. With 11 agents, adding enough judges can eventually bring peak accuracy to slightly above the simple majority, though it requires roughly an even split between judges and experts (or even slightly more judges). Further experiments show that if we simply add a judge, increasing the number of agents by one without reducing the number of experts, accuracy drops, so adding a judge is harmful, up until the point where the lower bound on competences guarantees the judge will be helpful (see Appendix).

In contrast to the uniform distribution, when drawing competences from the normal and the exponential distributions, things are a bit different. They show that a single judge can be productive. With normal distributions, when the mean is high but not extremely so (Figures 3b and 3e), adding a judge helps. When the mean is very high (0.8 and above), aggregating all agents as experts seems to be better than having a judge, for whom there is still a probability of being bad. But when agents are with a lower mean, having a judge seems to help, and this is true even for a mean of 0.5, in which there is a probability of 0.5 that the judge will be incompetent. This pattern appears for 5 agents, but, as in the uniform case, it is more pronounced for 11 agents.

For the exponential distributions (Figures 3c and 3f), this property is stronger – it is *always* beneficial, for our parameters, for agents to have one judge, and that improves over a simple majority. This is likely due to the fact that the small loss of accuracy from having one fewer experts is made up for by the ability of even a minimally competent judge (and all agents are competent in this distribution) to distinguish highly competent experts from less competent ones. Unlike in the uniform case, adding more judges (after a single one) is never helpful compared to a single judge (though sometimes two judges are better than simple majority).

These results imply that the division of labor in scientific conference is counter-intuitive: multiple layers above regular experts (e.g., SPC, AC) does not seem to be helpful. It seems better to have a flat hierarchy (i.e., fewer judges) and use simple majority, despite this idea being often frowned upon. We did not investigate the case where a judge is more competent than experts, but even then it is not clear more judges are better, as losing a top expert incurs a cost. Indeed, it is not at all clear that the best judge is the one with the highest competence, and we leave this open question to future research.



(a) 5 agents with competence from uniform distribution over $(\min, 1)$

(b) 5 agents with competence from truncated (in $(0, 1)$**) Gaussian competences** $\mathcal{N}(\textbf{mean}, 0.1)$

(c) 5 agents with competence from truncated (in $(0.5, 1)$**) exponential distribution**

(d) 11 agents with competence from uniform distribution over $(\min, 1)$

(e) 11 agents with competence from truncated (in $(0, 1)$**) Gaussian competences** $\mathcal{N}(\textbf{mean}, 0.1)$

(f) 11 agents with competence from truncated (in $(0.5, 1)$**) exponential distribution**

Figure 3: Accuracy partitioning a set of agents randomly into judges and experts with agent competences i.i.d. from the same distribution.

6 DISCUSSION AND FUTURE WORK

We consider a multi-level jury problem in which experts are given weights according to estimates from judges of their competence. We focus on settings where there is a small number of agents, so the classic asymptotic results from the literature do not apply, as well as cases where it is possible for agents to be incompetent (i.e., their chance of being correct might be less than 1/2).

We show that when judges are even minimally competent, we can guarantee that the weak order and sign of the weights assigned to experts will be correct. Additionally, we showed that if judges use the log-odds weighting and are reasonably accurate as a group, we will recover the optimal weighting function. Moreover, we show some cases where judges bring a meaningful benefit to the process. However, our results regarding how to divide a group of agents – a particularly relevant issue for scientific conferences – indicate that multiple judges may be unhelpful, and there are cases (e.g., uniform distributions) in which an additional expert is more valuable than a judge.

There are several interesting future directions. One is to reconsider the problem we have presented here when the weights given by experts must all be non-negative, or when it is required for each judge j that $\sum_{e \in E} w_{je} = 1$ (as required in Aziz et al. [2, 3] for the setting of peer evaluation). Another is to examine what happens when the hierarchy level is increased by adding an additional layers (as in large conferences, which have Area Chairs in charge of SPCs, in charge of PC members). At what point does it no longer become helpful (or begin to be helpful)? Can a guarantee of minimal quality of judges change the value proposition of having them? Further, it would be interesting to explore how this framework extends to non-binary information, as when PCs provide more detailed information to SPCs than just acceptance or rejection.

7 ACKNOWLEDGMENTS

Nicholas Mattei was supported by NSF Awards IIS-RI-2007955, IIS-III-2107505, IIS-RI-2134857, IIS-RI-2339880 and CNS-SCC-2427237 as well as the Tulane University Jurist Center for Artificial Intelligence and the Tulane University Center for Community-Engaged Artificial Intelligence. Ben Abramowitz was supported by the NSF under Grant #2127309 to the Computing Research Association for the CIFellows Project. Omer Lev was supported, in part, by NSF-BSF grant #2021659, and by Israel Science Fund (ISF) grants #1965/20 and #3152/20.

REFERENCES

[1] Ben Abramowitz and Nicholas Mattei. 2019. Flexible representative democracy: an introduction with binary issues. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence.* 3–10.

[2] Haris Aziz, Omer Lev, Nicholas Mattei, Jeffrey S. Rosenschein, and Toby Walsh. 2016. Strategyproof Peer Selection: Mechanisms, Analyses, and Experiments. In *Proceedings of the 30th Conference on Artificial Intelligence (AAAI)*. Phoenix, Arizona, 397–403.

[3] Haris Aziz, Omer Lev, Nicholas Mattei, Jeffrey S. Rosenschein, and Toby Walsh. 2019. Strategyproof peer selection using randomization, partitioning, and apportionment. *Artificial Intelligence (AIJ)* 275 (October 2019), 295–309. https://doi.org/10.1016/j.artint.2019.06.004

[4] Eyal Baharad, Jacob Goldberger, Moshe Koppel, and Shmuel Nitzan. 2012. Beyond Condorcet: Optimal aggregation rules using voting records. *Theory and decision* 72, 1 (2012), 113–130.

[5] Roy Baharad, Shmuel Nitzan, and Erel Segal-Halevi. 2022. One person, one weight: when is weighted voting democratic? *Social Choice and Welfare* (2022), 1–27.

[6] Itzhack Bar-Itzhack and Richard Harman. 2003. The effect of Sensor Failure on the Attitude and Rate Estimation of the MAP Spacecraft. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*. 5485.

[7] Ruben Becker, Gianlorenzo D'angelo, Esmaeil Delfaraz, and Hugo Gilbert. 2021. Unveiling the Truth in Liquid Democracy with Misinformed Voters. In *International Conference on Algorithmic Decision Theory*. Springer, 132–146.

[8] Ruth Ben-Yashar and Jacob Paroush. 2000. A nonasymptotic Condorcet jury theorem. *Social Choice and Welfare* 17, 2 (2000), 189–199.

[9] Daniel Berend and Aryeh Kontorovich. 2015. A finite sample analysis of the Naive Bayes classifier. *J. Mach. Learn. Res.* 16, 1 (2015), 1519–1545.

[10] Daniel Berend and Jacob Paroush. 1998. When is Condorcet's jury theorem valid? *Social Choice and Welfare* 15, 4 (1998), 481–488.

[11] Daniel Berend and Luba Sapir. 2003. Between the expert and majority rules. *Advances in Applied Probability* 35, 4 (2003), 941–960.

[12] Daren C Brabham. 2013. *Using crowdsourcing in government*. IBM Center for the Business of Government Washington, DC.

[13] Daren C Brabham. 2015. *Crowdsourcing in the public sector*. Georgetown University Press.

[14] Nicolo Cesa-Bianchi, Yoav Freund, David Haussler, David P Helmbold, Robert E Schapire, and Manfred K Warmuth. 1997. How to use expert advice. *Journal of the ACM (JACM)* 44, 3 (1997), 427–485.

[15] Marquis De Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. De l'Imprimerie Royale (Paris).

[16] Franz Dietrich and Kai Spiekermann. 2023. Jury Theorems. In *The Stanford Encyclopedia of Philosophy* (Spring 2023 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University.

[17] Scott L Feld and Bernard Grofman. 1984. The accuracy of group majority decisions in groups with added members. *Public Choice* 42, 3 (1984), 273–285.

[18] Bailey Flanigan, Paul Gölz, Anupam Gupta, Brett Hennig, and Ariel D. Procaccia. 2021. Fair algorithms for selecting citizens' assemblies. *Nature* 596 (2021), 548–552.

[19] Rupert Freeman, David Pennock, Chara Podimata, and Jennifer Wortman Vaughan. 2020. No-regret and incentive-compatible online learning. In *International Conference on Machine Learning*. PMLR, 3270–3279.

[20] Bernard Grofman. 1978. Judgmental competence of individuals and groups in a dichotomous choice situation: Is a majority of heads better than one? *Journal of Mathematical Sociology* 6, 1 (1978), 47–60.

[21] Bernard Grofman and Scott L Feld. 1983. Determining optimal weights for expert judgment. In *Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine, Conference on Political Economy*. JAI Press Greenwich, CT, 167–72.

[22] Bernard Grofman, Scott L Feld, and Guillermo Owen. 1984. Group size and the performance of a composite group majority: Statistical truths and empirical results. *Organizational Behavior and Human Performance* 33, 3 (1984), 350–359.

[23] Bernard Grofman, Guillermo Owen, and Scott L Feld. 1983. Thirteen theorems in search of the truth. *Theory and decision* 15, 3 (1983), 261–278.

[24] Drora Karotkin, Samuel Nitzal, and Jacob Paroush. 1988. The essential ranking of decision rules in small panels of experts. *Theory and Decision* 24 (1988), 253–268.

[25] Drora Karotkin and Jacob Paroush. 1994. Variability of decisional ability and the essential order of decision rules. *Journal of Economic Behavior & Organization* 23, 3 (1994), 343–354.

[26] Omer Lev, Nicholas Mattei, Paolo Turrini, and Stanislav Zhydkov. 2023. PeerNomination: A novel peer selection algorithm to handle strategic and noisy assessments. *Artificial Intelligence (AIJ)* 316 (March 2023), 103843.

[27] Malik Magdon-Ismail and Lirong Xia. 2018. A mathematical model for optimal decisions in a representative democracy. *Advances in Neural Information Processing Systems* 31 (2018).

[28] Shmuel Nitzan and Jacob Paroush. 1982. Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review* (1982), 289–297.

[29] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.

[30] Marcus Pivato and Arnold Soh. 2020. Weighted representative democracy. *Journal of Mathematical Economics* 88 (2020), 52–63.

[31] Manon Revel, Tao Lin, and Daniel Halpern. 2021. The Optimal Size of an Epistemic Congress. *arXiv preprint arXiv:2107.01042* (2021).

[32] Jan-Willem Romeijn and David Atkinson. 2011. Learning juror competence: A generalized Condorcet jury theorem. *Politics, Philosophy & Economics* 10, 3 (2011), 237–262.

[33] Robert E Schapire and Yoav Freund. 2013. Boosting: Foundations and algorithms. *Kybernetes* 42, 1 (2013), 164–166.

[34] Nihar B. Shah. 2022. Challenges, experiments, and computational solutions in peer review. *Commun. ACM* 65, 6 (2022), 76–87.

[35] Lloyd Shapley and Bernard Grofman. 1984. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice* 43, 3 (1984), 329–343.

[36] James Surowiecki. 2005. *The wisdom of crowds*. Anchor.

[37] Nassim Nicholas Taleb. 2007. *The Black Swan: The Impact of the Highly Improbable*. Random House.

[38] Alan Taylor and William Zwicker. 1992. A characterization of weighted voting. *Proceedings of the American mathematical society* 115, 4 (1992), 1089–1094.

[39] Zoi Terzopoulou. 2023. Voting with limited energy: A study of plurality and Borda. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 2085–2093.

[40] Chris Thornton, Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. 2013. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 847–855.

[41] Volodimir G Vovk. 1990. Aggregating strategies. *Proc. of Computational Learning Theory, 1990* (1990).

[42] Toby Walsh. 2014. The PeerRank Method for Peer Assessment. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI)*. Prague, Czech Republic, 909–914.

[43] Jens Witkowski and David C Parkes. 2012. Peer prediction without a common prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. 964–981.

[44] Yuzhe Zhang and Davide Grossi. 2022. Tracking Truth by Weighting Proxies in Liquid Democracy. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 1482–1490.

A TABLES

The following tables give the accuracy with competences drawn from various distributions given a number of judges *n* and number of experts *m*.

A.1 Uniformly Distributed competences

Table 4: Accuracy with judge and expert competences from uniform distribution over (0.1,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.549	0.550	0.572	0.574	0.591	0.591	0.607	0.607	0.621	0.621
1	0.526	0.534	0.538	0.544	0.546	0.549	0.550	0.552	0.550	0.550
2	0.546	0.563	0.574	0.582	0.587	0.590	0.592	0.593	0.596	0.597
3	0.559	0.579	0.593	0.603	0.607	0.611	0.615	0.621	0.623	0.622
4	0.571	0.593	0.611	0.618	0.627	0.631	0.636	0.641	0.642	0.643
5	0.577	0.603	0.620	0.633	0.641	0.649	0.653	0.657	0.661	0.664
6	0.583	0.616	0.635	0.645	0.655	0.663	0.669	0.671	0.677	0.679
7	0.591	0.624	0.643	0.660	0.667	0.675	0.679	0.685	0.691	0.694
8	0.598	0.631	0.653	0.667	0.681	0.685	0.694	0.700	0.704	0.706
9	0.603	0.638	0.662	0.677	0.692	0.700	0.706	0.712	0.714	0.718
10	0.608	0.646	0.669	0.688	0.701	0.710	0.716	0.721	0.726	0.728

Table 5: Accuracy with judge and expert competences from uniform distribution over (0.2,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.600	0.601	0.647	0.647	0.682	0.681	0.710	0.710	0.734	0.734
1	0.554	0.572	0.585	0.593	0.599	0.605	0.610	0.613	0.615	0.617
2	0.592	0.626	0.647	0.663	0.674	0.684	0.691	0.695	0.700	0.704
3	0.613	0.656	0.683	0.702	0.718	0.728	0.736	0.743	0.747	0.752
4	0.631	0.679	0.710	0.732	0.745	0.760	0.770	0.775	0.782	0.788
5	0.643	0.696	0.730	0.753	0.770	0.782	0.794	0.804	0.810	0.815
6	0.653	0.709	0.744	0.770	0.790	0.804	0.815	0.825	0.833	0.837
7	0.661	0.720	0.758	0.785	0.806	0.821	0.833	0.842	0.849	0.855
8	0.668	0.728	0.769	0.797	0.817	0.835	0.847	0.857	0.865	0.873
9	0.673	0.736	0.778	0.810	0.827	0.845	0.859	0.870	0.877	0.883
10	0.679	0.744	0.786	0.817	0.837	0.857	0.871	0.879	0.889	0.895

Table 6: Accuracy with judge and expert competences from uniform distribution over (0.3,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.651	0.649	0.717	0.718	0.762	0.763	0.800	0.798	0.829	0.827
1	0.588	0.623	0.645	0.661	0.673	0.681	0.687	0.692	0.696	0.702
2	0.640	0.694	0.728	0.752	0.770	0.784	0.794	0.801	0.808	0.815
3	0.665	0.728	0.770	0.800	0.819	0.835	0.846	0.855	0.863	0.869
4	0.678	0.749	0.792	0.823	0.846	0.865	0.877	0.887	0.897	0.903
5	0.688	0.760	0.808	0.841	0.864	0.883	0.897	0.909	0.917	0.923
6	0.694	0.768	0.819	0.851	0.877	0.895	0.909	0.921	0.931	0.938
7	0.698	0.774	0.825	0.859	0.884	0.903	0.919	0.931	0.940	0.946
8	0.700	0.778	0.829	0.865	0.890	0.909	0.923	0.936	0.946	0.953
9	0.702	0.782	0.831	0.867	0.893	0.913	0.928	0.940	0.949	0.957
10	0.705	0.784	0.835	0.869	0.895	0.915	0.931	0.942	0.952	0.960

Table 7: Accuracy with judge and expert competences from uniform distribution over (0.4,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.700	0.700	0.784	0.783	0.836	0.835	0.872	0.873	0.901	0.901
1	0.645	0.702	0.734	0.760	0.776	0.788	0.800	0.808	0.812	0.815
2	0.690	0.768	0.815	0.845	0.869	0.885	0.898	0.907	0.915	0.923
3	0.708	0.790	0.841	0.873	0.899	0.917	0.929	0.940	0.948	0.954
4	0.714	0.798	0.849	0.883	0.909	0.928	0.942	0.951	0.960	0.966
5	0.714	0.802	0.852	0.888	0.911	0.931	0.946	0.956	0.965	0.970
6	0.716	0.802	0.855	0.889	0.915	0.932	0.947	0.958	0.966	0.972
7	0.715	0.804	0.854	0.891	0.915	0.935	0.948	0.959	0.968	0.974
8	0.715	0.803	0.855	0.889	0.917	0.935	0.948	0.959	0.968	0.974
9	0.716	0.802	0.856	0.889	0.915	0.935	0.948	0.960	0.968	0.974
10	0.716	0.803	0.855	0.890	0.915	0.935	0.949	0.960	0.968	0.975

Table 8: Accuracy with judge and expert competences from uniform distribution over (0.5,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.750	0.749	0.843	0.843	0.897	0.897	0.929	0.929	0.950	0.950
1	0.750	0.831	0.880	0.913	0.936	0.951	0.964	0.973	0.980	0.984
2	0.751	0.833	0.881	0.914	0.937	0.953	0.964	0.974	0.980	0.984
3	0.748	0.833	0.880	0.914	0.937	0.952	0.964	0.972	0.980	0.984
4	0.748	0.832	0.882	0.914	0.937	0.952	0.964	0.974	0.980	0.984
5	0.747	0.833	0.882	0.915	0.937	0.953	0.965	0.972	0.980	0.984
6	0.748	0.832	0.881	0.914	0.937	0.952	0.964	0.974	0.980	0.985
7	0.750	0.833	0.881	0.913	0.937	0.954	0.964	0.974	0.980	0.984
8	0.748	0.834	0.881	0.915	0.937	0.952	0.964	0.974	0.980	0.985
9	0.751	0.832	0.881	0.914	0.937	0.953	0.965	0.974	0.980	0.984
10	0.748	0.833	0.881	0.915	0.937	0.954	0.966	0.974	0.979	0.984

Table 9: Accuracy with judge and expert competences from uniform distribution over (0.6,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.798	0.800	0.897	0.895	0.942	0.942	0.966	0.968	0.980	0.980
1	0.799	0.865	0.911	0.940	0.958	0.972	0.980	0.986	0.990	0.993
2	0.800	0.867	0.911	0.940	0.960	0.971	0.980	0.986	0.990	0.992
3	0.798	0.867	0.911	0.940	0.958	0.972	0.980	0.986	0.990	0.993
4	0.799	0.865	0.911	0.941	0.959	0.972	0.980	0.986	0.990	0.993
5	0.800	0.865	0.912	0.941	0.958	0.972	0.980	0.986	0.990	0.993
6	0.800	0.865	0.911	0.940	0.959	0.972	0.980	0.986	0.990	0.993
7	0.800	0.865	0.912	0.940	0.958	0.971	0.980	0.986	0.990	0.993
8	0.798	0.867	0.912	0.940	0.958	0.972	0.980	0.986	0.990	0.993
9	0.799	0.866	0.912	0.941	0.960	0.972	0.980	0.986	0.990	0.993
10	0.799	0.867	0.912	0.940	0.958	0.973	0.980	0.986	0.990	0.993

Table 10: Accuracy with judge and expert competences from uniform distribution over (0.7,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.849	0.849	0.940	0.938	0.973	0.973	0.988	0.988	0.994	0.994
1	0.850	0.899	0.942	0.965	0.978	0.987	0.991	0.994	0.996	0.998
2	0.850	0.899	0.942	0.966	0.979	0.986	0.992	0.995	0.996	0.998
3	0.849	0.901	0.943	0.965	0.978	0.987	0.991	0.994	0.997	0.998
4	0.851	0.897	0.942	0.965	0.978	0.987	0.991	0.994	0.996	0.998
5	0.851	0.899	0.943	0.966	0.978	0.987	0.992	0.995	0.996	0.998
6	0.849	0.899	0.944	0.966	0.978	0.986	0.992	0.994	0.996	0.998
7	0.849	0.899	0.942	0.966	0.978	0.987	0.992	0.995	0.996	0.998
8	0.849	0.901	0.942	0.966	0.978	0.986	0.992	0.995	0.996	0.998
9	0.851	0.901	0.942	0.965	0.978	0.987	0.991	0.995	0.997	0.998
10	0.850	0.898	0.943	0.966	0.978	0.986	0.992	0.994	0.996	0.998

Table 11: Accuracy with judge and expert competences from uniform distribution over (0.8,1.0)

$n\backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.900	0.900	0.972	0.972	0.992	0.992	0.997	0.997	0.999	0.999
1	0.901	0.933	0.972	0.984	0.992	0.996	0.998	0.999	1.000	1.000
2	0.899	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
3	0.901	0.933	0.971	0.984	0.992	0.996	0.998	0.999	0.999	1.000
4	0.899	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
5	0.901	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
6	0.899	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
7	0.900	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
8	0.899	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
9	0.899	0.931	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000
10	0.899	0.933	0.972	0.984	0.992	0.996	0.998	0.999	0.999	1.000

Table 12: Accuracy with judge and expert competences from uniform distribution over (0.9,1.0)

$n \backslash m$	1	2	3	4	5	6	7	8	9	10
0	0.949	0.950	0.993	0.992	0.999	0.999	1.000	1.000	1.000	1.000
1	0.949	0.965	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
2	0.948	0.966	0.993	0.996	0.999	0.999	1.000	1.000	1.000	1.000
3	0.948	0.966	0.993	0.996	0.999	1.000	1.000	1.000	1.000	1.000
4	0.949	0.966	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
5	0.949	0.966	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
6	0.950	0.966	0.993	0.996	0.999	1.000	1.000	1.000	1.000	1.000
7	0.950	0.966	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
8	0.949	0.966	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
9	0.948	0.965	0.992	0.996	0.999	0.999	1.000	1.000	1.000	1.000
10	0.948	0.966	0.992	0.996	0.999	1.000	1.000	1.000	1.000	1.000

A.2 Normally Distributed competences

Table 13: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.5,0.1)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.581	0.578	0.621	0.618	0.647	0.648	0.673	0.669	0.693	0.692
1	0.580	0.616	0.639	0.658	0.673	0.691	0.705	0.720	0.732	0.741
2	0.580	0.612	0.640	0.660	0.680	0.690	0.705	0.719	0.729	0.743
3	0.581	0.612	0.640	0.659	0.676	0.691	0.707	0.718	0.730	0.742
4	0.581	0.614	0.638	0.658	0.673	0.692	0.707	0.719	0.727	0.741
5	0.581	0.616	0.635	0.658	0.674	0.692	0.706	0.718	0.729	0.743
6	0.580	0.612	0.635	0.660	0.675	0.692	0.706	0.717	0.731	0.742
7	0.579	0.613	0.639	0.659	0.678	0.692	0.707	0.718	0.730	0.741
8	0.581	0.615	0.639	0.659	0.676	0.689	0.708	0.718	0.732	0.741
9	0.582	0.613	0.639	0.659	0.675	0.692	0.707	0.719	0.733	0.741
10	0.579	0.612	0.641	0.662	0.676	0.692	0.705	0.718	0.731	0.743

Table 14: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.5,0.4)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.719	0.720	0.807	0.806	0.861	0.862	0.897	0.899	0.923	0.923
1	0.718	0.798	0.850	0.882	0.908	0.927	0.941	0.953	0.964	0.971
2	0.718	0.798	0.848	0.883	0.909	0.928	0.942	0.953	0.963	0.970
3	0.720	0.797	0.850	0.882	0.908	0.927	0.942	0.953	0.964	0.969
4	0.721	0.800	0.848	0.884	0.908	0.927	0.942	0.955	0.964	0.971
5	0.721	0.801	0.850	0.885	0.907	0.926	0.943	0.954	0.963	0.970
6	0.721	0.799	0.850	0.884	0.910	0.928	0.943	0.954	0.965	0.972
7	0.719	0.800	0.851	0.883	0.908	0.927	0.943	0.955	0.963	0.971
8	0.721	0.801	0.848	0.883	0.910	0.927	0.943	0.955	0.963	0.971
9	0.721	0.798	0.846	0.883	0.909	0.927	0.942	0.954	0.963	0.971
10	0.722	0.800	0.850	0.883	0.909	0.928	0.943	0.954	0.963	0.971

Table 15: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.6,0.1)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.629	0.629	0.689	0.689	0.732	0.732	0.765	0.764	0.791	0.789
1	0.628	0.673	0.708	0.737	0.762	0.778	0.799	0.815	0.829	0.842
2	0.630	0.675	0.709	0.735	0.760	0.780	0.798	0.814	0.830	0.840
3	0.629	0.674	0.707	0.737	0.761	0.782	0.797	0.814	0.830	0.842
4	0.629	0.674	0.712	0.738	0.762	0.780	0.798	0.815	0.827	0.839
5	0.630	0.673	0.710	0.736	0.761	0.779	0.797	0.815	0.829	0.840
6	0.629	0.675	0.711	0.737	0.761	0.783	0.799	0.814	0.829	0.839
7	0.628	0.674	0.709	0.736	0.760	0.782	0.796	0.815	0.828	0.842
8	0.632	0.672	0.711	0.738	0.757	0.780	0.798	0.814	0.830	0.840
9	0.630	0.677	0.709	0.736	0.760	0.781	0.797	0.817	0.828	0.843
10	0.630	0.673	0.706	0.736	0.761	0.782	0.798	0.814	0.827	0.840

Table 16: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.6,0.4)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.732	0.731	0.825	0.820	0.878	0.875	0.911	0.911	0.936	0.935
1	0.732	0.811	0.861	0.896	0.920	0.939	0.952	0.962	0.969	0.976
2	0.731	0.812	0.861	0.895	0.919	0.937	0.951	0.963	0.970	0.977
3	0.729	0.812	0.861	0.895	0.922	0.939	0.953	0.962	0.970	0.976
4	0.730	0.813	0.862	0.896	0.919	0.938	0.952	0.962	0.969	0.977
5	0.733	0.812	0.861	0.895	0.920	0.939	0.952	0.963	0.971	0.977
6	0.731	0.811	0.861	0.897	0.920	0.939	0.952	0.963	0.971	0.976
7	0.733	0.812	0.861	0.895	0.920	0.938	0.950	0.962	0.970	0.976
8	0.732	0.814	0.861	0.895	0.919	0.937	0.953	0.963	0.971	0.977
9	0.731	0.811	0.860	0.895	0.919	0.938	0.952	0.963	0.970	0.977
10	0.732	0.811	0.859	0.897	0.921	0.938	0.952	0.963	0.970	0.976

Table 17: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}$(0.7,0.1) over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.706	0.706	0.792	0.792	0.843	0.845	0.881	0.880	0.908	0.907
1	0.704	0.758	0.805	0.838	0.866	0.887	0.904	0.917	0.931	0.941
2	0.703	0.761	0.803	0.840	0.864	0.886	0.904	0.916	0.930	0.940
3	0.705	0.758	0.806	0.839	0.866	0.885	0.903	0.916	0.930	0.940
4	0.704	0.762	0.805	0.837	0.865	0.886	0.903	0.919	0.930	0.940
5	0.704	0.761	0.804	0.838	0.864	0.885	0.905	0.919	0.931	0.940
6	0.702	0.755	0.804	0.837	0.864	0.885	0.903	0.918	0.930	0.939
7	0.706	0.758	0.805	0.837	0.865	0.886	0.902	0.917	0.930	0.939
8	0.706	0.757	0.805	0.838	0.866	0.886	0.903	0.917	0.929	0.942
9	0.704	0.761	0.805	0.836	0.865	0.888	0.901	0.918	0.930	0.940
10	0.704	0.758	0.803	0.838	0.866	0.887	0.901	0.917	0.930	0.940

Table 18: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}$(0.7,0.4) over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.742	0.744	0.837	0.835	0.889	0.890	0.924	0.923	0.945	0.946
1	0.744	0.823	0.873	0.904	0.929	0.948	0.959	0.968	0.976	0.981
2	0.743	0.826	0.873	0.906	0.930	0.946	0.960	0.969	0.976	0.982
3	0.743	0.824	0.874	0.906	0.929	0.948	0.959	0.969	0.975	0.981
4	0.745	0.827	0.873	0.907	0.931	0.947	0.961	0.968	0.977	0.981
5	0.743	0.826	0.871	0.905	0.928	0.947	0.960	0.969	0.977	0.982
6	0.743	0.827	0.874	0.906	0.929	0.946	0.958	0.969	0.976	0.981
7	0.745	0.825	0.874	0.906	0.929	0.946	0.959	0.969	0.977	0.982
8	0.744	0.824	0.873	0.906	0.930	0.948	0.959	0.969	0.976	0.982
9	0.748	0.823	0.875	0.906	0.929	0.948	0.960	0.968	0.976	0.981
10	0.741	0.824	0.873	0.907	0.929	0.948	0.960	0.968	0.977	0.982

Table 19: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.8,0.1)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.796	0.796	0.890	0.892	0.937	0.940	0.964	0.964	0.979	0.979
1	0.793	0.848	0.897	0.926	0.947	0.962	0.972	0.980	0.985	0.989
2	0.794	0.849	0.899	0.927	0.948	0.962	0.973	0.979	0.985	0.989
3	0.795	0.849	0.898	0.927	0.947	0.963	0.972	0.979	0.984	0.989
4	0.793	0.847	0.898	0.926	0.948	0.962	0.973	0.980	0.985	0.989
5	0.795	0.847	0.898	0.927	0.946	0.961	0.971	0.980	0.985	0.989
6	0.796	0.848	0.897	0.926	0.949	0.962	0.972	0.980	0.985	0.990
7	0.794	0.848	0.898	0.928	0.947	0.962	0.972	0.980	0.986	0.989
8	0.795	0.849	0.897	0.929	0.948	0.962	0.971	0.979	0.985	0.989
9	0.795	0.847	0.898	0.928	0.949	0.961	0.972	0.980	0.985	0.989
10	0.794	0.848	0.897	0.927	0.947	0.963	0.972	0.980	0.985	0.989

Table 20: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.8,0.4)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.756	0.756	0.849	0.851	0.902	0.903	0.935	0.935	0.956	0.956
1	0.759	0.836	0.887	0.916	0.940	0.954	0.966	0.974	0.981	0.985
2	0.757	0.838	0.886	0.916	0.939	0.955	0.967	0.975	0.980	0.986
3	0.757	0.837	0.885	0.916	0.940	0.955	0.966	0.975	0.981	0.986
4	0.758	0.838	0.887	0.916	0.940	0.956	0.966	0.975	0.981	0.986
5	0.755	0.836	0.884	0.918	0.940	0.954	0.967	0.975	0.981	0.986
6	0.756	0.837	0.885	0.917	0.939	0.955	0.967	0.974	0.981	0.985
7	0.756	0.836	0.885	0.918	0.939	0.954	0.966	0.975	0.981	0.985
8	0.757	0.838	0.885	0.918	0.938	0.955	0.966	0.974	0.981	0.985
9	0.757	0.838	0.884	0.917	0.939	0.955	0.966	0.974	0.982	0.986
10	0.755	0.838	0.886	0.917	0.939	0.955	0.966	0.975	0.981	0.985

Table 21: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.9,0.1)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.871	0.870	0.954	0.954	0.983	0.982	0.993	0.993	0.997	0.997
1	0.871	0.916	0.956	0.975	0.986	0.991	0.995	0.997	0.998	1.0
2	0.872	0.915	0.957	0.975	0.985	0.992	0.995	0.997	0.998	1.0
3	0.870	0.917	0.957	0.975	0.985	0.992	0.995	0.997	0.998	1.0
4	0.871	0.915	0.956	0.975	0.985	0.991	0.995	0.997	0.998	1.0
5	0.869	0.916	0.956	0.974	0.985	0.992	0.995	0.997	0.998	1.0
6	0.870	0.916	0.957	0.975	0.985	0.991	0.994	0.997	0.998	1.0
7	0.871	0.915	0.957	0.976	0.986	0.992	0.995	0.997	0.998	1.0
8	0.871	0.915	0.956	0.974	0.986	0.992	0.995	0.997	0.998	1.0
9	0.869	0.914	0.956	0.975	0.985	0.991	0.995	0.997	0.998	1.0
10	0.871	0.914	0.957	0.975	0.986	0.991	0.994	0.997	0.998	1.0

Table 22: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.9{,}0.4)$ over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.768	0.767	0.866	0.865	0.914	0.913	0.946	0.946	0.964	0.964
1	0.766	0.850	0.894	0.926	0.947	0.961	0.972	0.980	0.985	0.989
2	0.768	0.847	0.895	0.926	0.947	0.962	0.972	0.979	0.985	0.990
3	0.767	0.850	0.896	0.926	0.946	0.961	0.972	0.980	0.985	0.989
4	0.768	0.849	0.896	0.927	0.947	0.962	0.972	0.979	0.985	0.989
5	0.769	0.847	0.896	0.925	0.947	0.962	0.973	0.979	0.985	0.989
6	0.767	0.848	0.897	0.925	0.947	0.961	0.972	0.980	0.986	0.989
7	0.767	0.850	0.895	0.928	0.948	0.962	0.973	0.979	0.985	0.989
8	0.769	0.849	0.895	0.927	0.947	0.962	0.972	0.981	0.985	0.990
9	0.768	0.850	0.896	0.927	0.948	0.963	0.973	0.980	0.985	0.989
10	0.769	0.847	0.895	0.925	0.947	0.961	0.972	0.980	0.985	0.988

Table 23: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.5,0.1)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.499	0.501	0.500	0.502	0.499	0.499	0.502	0.501	0.501	0.501
1	0.498	0.502	0.500	0.499	0.501	0.501	0.499	0.500	0.502	0.501
2	0.500	0.502	0.497	0.500	0.500	0.501	0.500	0.503	0.500	0.498
3	0.502	0.502	0.502	0.499	0.503	0.502	0.499	0.499	0.500	0.496
4	0.501	0.500	0.498	0.498	0.499	0.502	0.499	0.501	0.497	0.499
5	0.498	0.503	0.500	0.499	0.499	0.499	0.500	0.498	0.501	0.500
6	0.501	0.501	0.500	0.498	0.501	0.499	0.501	0.499	0.501	0.500
7	0.499	0.502	0.500	0.499	0.501	0.498	0.501	0.497	0.501	0.500
8	0.498	0.499	0.501	0.500	0.499	0.498	0.499	0.498	0.501	0.502
9	0.500	0.498	0.499	0.499	0.500	0.499	0.499	0.501	0.500	0.502
10	0.496	0.500	0.500	0.498	0.500	0.501	0.500	0.498	0.499	0.502

Table 24: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.5,0.4)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.502	0.501	0.499	0.498	0.497	0.499	0.496	0.503	0.500	0.498
1	0.498	0.498	0.502	0.499	0.498	0.497	0.498	0.498	0.499	0.499
2	0.500	0.499	0.498	0.498	0.497	0.499	0.497	0.499	0.501	0.498
3	0.501	0.502	0.500	0.499	0.500	0.499	0.499	0.497	0.498	0.500
4	0.499	0.501	0.502	0.497	0.499	0.499	0.499	0.498	0.501	0.500
5	0.500	0.499	0.498	0.500	0.498	0.501	0.498	0.496	0.499	0.499
6	0.500	0.500	0.498	0.499	0.499	0.495	0.499	0.498	0.499	0.502
7	0.500	0.498	0.498	0.501	0.500	0.498	0.498	0.500	0.499	0.502
8	0.498	0.500	0.500	0.501	0.497	0.499	0.500	0.498	0.499	0.497
9	0.500	0.497	0.497	0.496	0.502	0.496	0.501	0.497	0.500	0.497
10	0.498	0.498	0.498	0.499	0.501	0.502	0.497	0.498	0.499	0.496

Table 25: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.6,0.1)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.600	0.599	0.651	0.649	0.683	0.684	0.708	0.710	0.735	0.733
1	0.582	0.609	0.631	0.653	0.666	0.681	0.694	0.701	0.712	0.721
2	0.598	0.636	0.668	0.687	0.706	0.721	0.737	0.749	0.760	0.771
3	0.605	0.646	0.677	0.703	0.722	0.739	0.756	0.771	0.785	0.799
4	0.610	0.652	0.687	0.712	0.734	0.751	0.768	0.785	0.797	0.808
5	0.614	0.658	0.690	0.717	0.740	0.758	0.774	0.788	0.803	0.815
6	0.616	0.658	0.690	0.718	0.742	0.759	0.779	0.792	0.807	0.819
7	0.617	0.663	0.695	0.721	0.742	0.761	0.779	0.795	0.809	0.822
8	0.618	0.660	0.697	0.722	0.744	0.762	0.781	0.797	0.810	0.822
9	0.617	0.663	0.692	0.722	0.746	0.764	0.783	0.796	0.808	0.822
10	0.620	0.662	0.697	0.724	0.746	0.764	0.778	0.795	0.812	0.825

Table 26: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.6,0.4)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.541	0.538	0.562	0.566	0.574	0.578	0.590	0.590	0.599	0.603
1	0.532	0.543	0.547	0.552	0.553	0.556	0.558	0.563	0.563	0.565
2	0.543	0.553	0.566	0.573	0.576	0.579	0.580	0.582	0.586	0.589
3	0.551	0.567	0.576	0.586	0.594	0.595	0.599	0.598	0.606	0.603
4	0.554	0.577	0.590	0.596	0.602	0.611	0.611	0.617	0.617	0.617
5	0.561	0.586	0.598	0.611	0.616	0.621	0.623	0.632	0.634	0.632
6	0.569	0.592	0.603	0.617	0.628	0.632	0.637	0.638	0.643	0.645
7	0.573	0.599	0.616	0.626	0.635	0.643	0.646	0.649	0.655	0.657
8	0.577	0.605	0.625	0.635	0.645	0.650	0.657	0.661	0.664	0.664
9	0.579	0.612	0.630	0.646	0.652	0.661	0.662	0.668	0.674	0.677
10	0.586	0.617	0.635	0.647	0.661	0.669	0.672	0.678	0.680	0.683

Table 27: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.7,0.1)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.701	0.697	0.785	0.785	0.838	0.837	0.873	0.873	0.902	0.901
1	0.690	0.744	0.790	0.820	0.845	0.866	0.883	0.897	0.908	0.917
2	0.699	0.753	0.797	0.832	0.859	0.880	0.901	0.913	0.925	0.934
3	0.703	0.755	0.802	0.837	0.861	0.881	0.900	0.915	0.927	0.937
4	0.702	0.755	0.801	0.836	0.860	0.884	0.900	0.916	0.927	0.937
5	0.699	0.752	0.800	0.838	0.861	0.883	0.901	0.916	0.926	0.938
6	0.703	0.754	0.803	0.836	0.861	0.882	0.899	0.915	0.926	0.937
7	0.700	0.755	0.800	0.836	0.860	0.884	0.900	0.915	0.927	0.937
8	0.702	0.755	0.803	0.836	0.860	0.883	0.901	0.914	0.927	0.937
9	0.701	0.758	0.802	0.837	0.863	0.884	0.898	0.916	0.927	0.938
10	0.701	0.754	0.802	0.835	0.861	0.883	0.900	0.915	0.927	0.938

Table 28: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.7,0.4)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.581	0.583	0.622	0.621	0.654	0.653	0.675	0.677	0.696	0.694
1	0.559	0.582	0.593	0.604	0.611	0.617	0.622	0.622	0.627	0.625
2	0.581	0.612	0.628	0.640	0.649	0.655	0.663	0.666	0.666	0.668
3	0.600	0.632	0.651	0.668	0.677	0.683	0.691	0.696	0.699	0.703
4	0.611	0.647	0.674	0.689	0.701	0.709	0.718	0.724	0.726	0.731
5	0.618	0.660	0.690	0.707	0.720	0.730	0.738	0.744	0.749	0.754
6	0.630	0.676	0.702	0.724	0.738	0.748	0.757	0.762	0.766	0.770
7	0.635	0.685	0.715	0.734	0.755	0.761	0.772	0.778	0.784	0.786
8	0.645	0.697	0.728	0.748	0.765	0.777	0.785	0.793	0.799	0.804
9	0.651	0.703	0.737	0.761	0.777	0.791	0.796	0.805	0.811	0.816
10	0.657	0.712	0.748	0.769	0.787	0.800	0.810	0.813	0.823	0.827

Table 29: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.8,0.1)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.795	0.795	0.891	0.892	0.937	0.938	0.964	0.964	0.979	0.978
1	0.793	0.846	0.896	0.925	0.946	0.962	0.972	0.978	0.983	0.988
2	0.794	0.848	0.900	0.928	0.948	0.962	0.972	0.980	0.985	0.988
3	0.795	0.847	0.897	0.927	0.947	0.962	0.972	0.980	0.985	0.989
4	0.796	0.848	0.898	0.927	0.947	0.962	0.973	0.979	0.985	0.989
5	0.794	0.849	0.899	0.928	0.948	0.961	0.972	0.979	0.985	0.988
6	0.794	0.848	0.898	0.928	0.949	0.962	0.972	0.980	0.985	0.989
7	0.791	0.849	0.897	0.927	0.947	0.962	0.971	0.979	0.985	0.989
8	0.793	0.846	0.898	0.927	0.948	0.961	0.972	0.979	0.985	0.989
9	0.794	0.848	0.898	0.928	0.947	0.962	0.972	0.979	0.985	0.989
10	0.796	0.847	0.898	0.927	0.949	0.962	0.972	0.980	0.984	0.989

Table 30: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.8,0.4)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.621	0.620	0.677	0.678	0.721	0.717	0.750	0.752	0.776	0.778
1	0.595	0.622	0.641	0.657	0.663	0.672	0.679	0.683	0.682	0.686
2	0.621	0.663	0.686	0.705	0.717	0.725	0.737	0.741	0.741	0.746
3	0.640	0.691	0.719	0.739	0.754	0.766	0.770	0.781	0.783	0.789
4	0.654	0.709	0.743	0.766	0.783	0.795	0.804	0.810	0.817	0.818
5	0.670	0.728	0.763	0.788	0.804	0.817	0.826	0.834	0.839	0.843
6	0.677	0.741	0.778	0.804	0.824	0.835	0.845	0.853	0.858	0.867
7	0.688	0.752	0.791	0.817	0.840	0.851	0.862	0.870	0.875	0.882
8	0.693	0.762	0.802	0.829	0.850	0.863	0.876	0.885	0.890	0.895
9	0.700	0.767	0.808	0.838	0.861	0.876	0.888	0.894	0.900	0.906
10	0.705	0.774	0.815	0.848	0.868	0.886	0.895	0.905	0.911	0.917

Table 31: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.9,0.1)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.870	0.871	0.955	0.953	0.983	0.982	0.993	0.993	0.997	0.997
1	0.872	0.914	0.956	0.975	0.985	0.992	0.995	0.997	0.998	1.0
2	0.871	0.916	0.957	0.975	0.985	0.992	0.995	0.997	0.998	1.0
3	0.871	0.915	0.956	0.975	0.985	0.991	0.995	0.997	0.998	1.0
4	0.871	0.916	0.957	0.975	0.986	0.992	0.995	0.997	0.998	1.0
5	0.871	0.915	0.957	0.974	0.985	0.992	0.995	0.997	0.998	1.0
6	0.871	0.917	0.957	0.975	0.985	0.991	0.995	0.997	0.998	1.0
7	0.869	0.913	0.957	0.974	0.985	0.992	0.995	0.997	1.0	1.0
8	0.869	0.916	0.956	0.974	0.985	0.991	0.995	0.997	0.998	1.0
9	0.871	0.914	0.957	0.975	0.986	0.992	0.995	0.997	0.998	1.0
10	0.872	0.915	0.957	0.974	0.985	0.992	0.995	0.997	0.998	1.0

Table 32: Accuracy with judge and expert competences from truncated Gaussian distribution $\mathcal{N}(0.9,0.4)$ over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.658	0.659	0.728	0.729	0.777	0.776	0.813	0.813	0.840	0.838
1	0.621	0.663	0.689	0.703	0.716	0.724	0.728	0.734	0.739	0.743
2	0.656	0.710	0.740	0.763	0.777	0.789	0.795	0.802	0.805	0.810
3	0.677	0.741	0.774	0.801	0.817	0.831	0.839	0.845	0.851	0.855
4	0.696	0.761	0.802	0.827	0.842	0.861	0.868	0.876	0.884	0.885
5	0.706	0.774	0.816	0.847	0.865	0.879	0.892	0.899	0.905	0.909
6	0.718	0.788	0.830	0.860	0.880	0.897	0.907	0.916	0.922	0.924
7	0.722	0.797	0.841	0.873	0.893	0.907	0.918	0.927	0.934	0.938
8	0.725	0.803	0.850	0.878	0.903	0.918	0.928	0.937	0.942	0.947
9	0.729	0.809	0.853	0.888	0.908	0.922	0.935	0.944	0.952	0.956
10	0.736	0.812	0.859	0.893	0.913	0.929	0.942	0.950	0.957	0.960

A.3 Exponentially Distributed competences

Table 33: Accuracy with judge and expert competences from exponential distribution with scale 1 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.711	0.710	0.795	0.795	0.847	0.848	0.886	0.885	0.912	0.912
1	0.708	0.789	0.837	0.874	0.899	0.920	0.937	0.949	0.957	0.966
2	0.707	0.794	0.840	0.873	0.903	0.920	0.937	0.949	0.959	0.966
3	0.710	0.788	0.839	0.876	0.901	0.921	0.936	0.949	0.958	0.967
4	0.710	0.788	0.841	0.873	0.900	0.921	0.937	0.947	0.959	0.966
5	0.708	0.792	0.839	0.876	0.902	0.921	0.937	0.949	0.959	0.966
6	0.709	0.790	0.840	0.874	0.901	0.921	0.936	0.949	0.958	0.966
7	0.710	0.786	0.839	0.873	0.900	0.921	0.937	0.949	0.960	0.967
8	0.712	0.789	0.838	0.874	0.902	0.922	0.936	0.948	0.960	0.966
9	0.711	0.789	0.840	0.876	0.902	0.923	0.936	0.950	0.958	0.967
10	0.708	0.789	0.840	0.877	0.902	0.922	0.936	0.949	0.958	0.966

Table 34: Accuracy with judge and expert competences from exponential distribution with scale 2 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.672	0.672	0.750	0.747	0.797	0.799	0.834	0.835	0.862	0.865
1	0.674	0.745	0.795	0.831	0.857	0.881	0.897	0.913	0.925	0.935
2	0.672	0.744	0.795	0.829	0.858	0.882	0.898	0.912	0.926	0.936
3	0.675	0.744	0.796	0.829	0.858	0.879	0.899	0.913	0.925	0.936
4	0.671	0.747	0.791	0.830	0.858	0.880	0.897	0.913	0.924	0.938
5	0.672	0.746	0.794	0.831	0.857	0.879	0.897	0.912	0.925	0.936
6	0.672	0.746	0.791	0.829	0.858	0.880	0.898	0.913	0.926	0.935
7	0.674	0.742	0.793	0.831	0.857	0.880	0.896	0.913	0.925	0.936
8	0.672	0.745	0.792	0.829	0.858	0.880	0.897	0.913	0.925	0.937
9	0.671	0.747	0.796	0.829	0.858	0.881	0.898	0.913	0.925	0.936
10	0.674	0.745	0.795	0.832	0.857	0.878	0.897	0.912	0.927	0.936

Table 35: Accuracy with judge and expert competences from exponential distribution with scale 3 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.640	0.641	0.705	0.706	0.751	0.751	0.785	0.783	0.813	0.811
1	0.639	0.705	0.750	0.786	0.813	0.833	0.853	0.870	0.882	0.895
2	0.636	0.704	0.749	0.783	0.811	0.834	0.852	0.870	0.883	0.896
3	0.642	0.706	0.747	0.783	0.813	0.836	0.851	0.870	0.884	0.896
4	0.642	0.706	0.751	0.785	0.811	0.833	0.852	0.870	0.885	0.896
5	0.639	0.703	0.750	0.784	0.811	0.835	0.852	0.870	0.885	0.896
6	0.640	0.708	0.749	0.783	0.810	0.835	0.853	0.869	0.886	0.897
7	0.642	0.705	0.750	0.785	0.810	0.832	0.851	0.870	0.884	0.897
8	0.642	0.702	0.748	0.786	0.810	0.831	0.853	0.868	0.883	0.898
9	0.639	0.704	0.748	0.783	0.811	0.833	0.855	0.870	0.884	0.896
10	0.639	0.704	0.750	0.783	0.811	0.834	0.853	0.871	0.883	0.896

Table 36: Accuracy with judge and expert competences from exponential distribution with scale 4 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.618	0.617	0.668	0.673	0.709	0.713	0.741	0.740	0.766	0.765
1	0.617	0.670	0.712	0.741	0.768	0.790	0.808	0.826	0.839	0.853
2	0.617	0.672	0.710	0.742	0.769	0.789	0.806	0.826	0.838	0.850
3	0.613	0.670	0.710	0.741	0.766	0.787	0.807	0.824	0.840	0.853
4	0.618	0.671	0.708	0.740	0.767	0.788	0.809	0.824	0.841	0.853
5	0.615	0.672	0.709	0.743	0.766	0.788	0.810	0.825	0.839	0.853
6	0.615	0.670	0.710	0.741	0.766	0.789	0.807	0.824	0.838	0.853
7	0.614	0.671	0.710	0.743	0.768	0.792	0.808	0.825	0.842	0.850
8	0.616	0.669	0.709	0.740	0.763	0.790	0.810	0.822	0.839	0.852
9	0.614	0.669	0.711	0.740	0.765	0.787	0.808	0.825	0.838	0.853
10	0.617	0.672	0.711	0.742	0.769	0.788	0.809	0.827	0.838	0.852

Table 37: Accuracy with judge and expert competences from exponential distribution with scale 5 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.597	0.596	0.645	0.644	0.678	0.677	0.703	0.705	0.729	0.727
1	0.599	0.646	0.681	0.704	0.733	0.748	0.767	0.782	0.797	0.810
2	0.594	0.642	0.681	0.708	0.733	0.750	0.764	0.784	0.797	0.811
3	0.596	0.644	0.679	0.707	0.727	0.751	0.767	0.785	0.799	0.810
4	0.595	0.646	0.679	0.704	0.728	0.751	0.769	0.783	0.794	0.808
5	0.598	0.642	0.679	0.703	0.728	0.750	0.769	0.780	0.797	0.810
6	0.598	0.646	0.679	0.705	0.726	0.750	0.768	0.781	0.795	0.809
7	0.601	0.642	0.681	0.704	0.730	0.750	0.768	0.783	0.799	0.808
8	0.600	0.642	0.679	0.705	0.728	0.747	0.766	0.783	0.797	0.811
9	0.598	0.644	0.679	0.705	0.728	0.749	0.769	0.784	0.798	0.811
10	0.597	0.643	0.679	0.706	0.731	0.748	0.766	0.781	0.796	0.809

Table 38: Accuracy with judge and expert competences from exponential distribution with scale 6 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.584	0.582	0.621	0.621	0.652	0.651	0.675	0.676	0.697	0.697
1	0.583	0.623	0.652	0.679	0.699	0.717	0.732	0.747	0.759	0.773
2	0.586	0.626	0.656	0.679	0.699	0.715	0.733	0.744	0.760	0.771
3	0.582	0.625	0.653	0.676	0.697	0.716	0.730	0.744	0.761	0.772
4	0.581	0.624	0.654	0.678	0.696	0.714	0.733	0.747	0.761	0.774
5	0.581	0.621	0.652	0.676	0.699	0.717	0.731	0.746	0.758	0.773
6	0.584	0.626	0.653	0.677	0.696	0.717	0.731	0.747	0.757	0.771
7	0.583	0.623	0.653	0.678	0.695	0.716	0.733	0.747	0.760	0.773
8	0.583	0.621	0.652	0.678	0.699	0.718	0.731	0.747	0.759	0.773
9	0.584	0.625	0.653	0.675	0.699	0.716	0.732	0.748	0.760	0.772
10	0.584	0.621	0.653	0.679	0.697	0.716	0.732	0.745	0.761	0.772

Table 39: Accuracy with judge and expert competences from exponential distribution with scale 7 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.574	0.570	0.608	0.607	0.634	0.635	0.652	0.654	0.672	0.671
1	0.572	0.606	0.634	0.654	0.672	0.689	0.702	0.717	0.727	0.740
2	0.571	0.606	0.632	0.655	0.675	0.689	0.703	0.718	0.730	0.742
3	0.572	0.610	0.634	0.654	0.672	0.688	0.702	0.715	0.727	0.742
4	0.574	0.607	0.636	0.654	0.673	0.689	0.701	0.716	0.728	0.741
5	0.572	0.605	0.636	0.656	0.672	0.690	0.702	0.716	0.730	0.741
6	0.571	0.606	0.635	0.655	0.671	0.691	0.703	0.716	0.730	0.737
7	0.573	0.607	0.632	0.655	0.677	0.690	0.703	0.716	0.728	0.739
8	0.576	0.607	0.631	0.655	0.673	0.689	0.705	0.716	0.728	0.738
9	0.569	0.606	0.634	0.657	0.672	0.686	0.703	0.718	0.728	0.738
10	0.570	0.608	0.634	0.656	0.671	0.688	0.703	0.716	0.727	0.741

Table 40: Accuracy with judge and expert competences from exponential distribution with scale 8 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.566	0.563	0.593	0.596	0.615	0.615	0.639	0.634	0.653	0.650
1	0.564	0.594	0.620	0.636	0.653	0.668	0.679	0.692	0.704	0.711
2	0.564	0.593	0.616	0.635	0.651	0.668	0.678	0.694	0.706	0.711
3	0.564	0.594	0.616	0.635	0.653	0.666	0.681	0.691	0.701	0.713
4	0.564	0.594	0.617	0.635	0.654	0.670	0.682	0.692	0.702	0.715
5	0.567	0.592	0.618	0.637	0.650	0.665	0.682	0.691	0.704	0.714
6	0.564	0.592	0.616	0.638	0.651	0.669	0.682	0.692	0.701	0.712
7	0.563	0.595	0.617	0.637	0.652	0.668	0.685	0.692	0.703	0.713
8	0.561	0.596	0.617	0.635	0.651	0.667	0.681	0.691	0.705	0.714
9	0.562	0.593	0.619	0.636	0.651	0.666	0.680	0.691	0.705	0.714
10	0.564	0.594	0.617	0.636	0.651	0.668	0.679	0.693	0.699	0.715

Table 41: Accuracy with judge and expert competences from exponential distribution with scale 9 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.557	0.557	0.585	0.583	0.607	0.602	0.620	0.619	0.636	0.637
1	0.558	0.584	0.603	0.623	0.640	0.649	0.663	0.671	0.681	0.693
2	0.552	0.583	0.602	0.622	0.636	0.650	0.663	0.669	0.684	0.693
3	0.559	0.586	0.604	0.620	0.638	0.647	0.662	0.672	0.684	0.693
4	0.554	0.585	0.604	0.623	0.636	0.649	0.662	0.672	0.685	0.692
5	0.558	0.582	0.605	0.624	0.633	0.650	0.662	0.672	0.683	0.689
6	0.556	0.583	0.610	0.624	0.638	0.649	0.662	0.669	0.683	0.688
7	0.554	0.584	0.604	0.623	0.637	0.649	0.665	0.672	0.681	0.692
8	0.556	0.586	0.603	0.622	0.640	0.649	0.662	0.670	0.684	0.694
9	0.557	0.584	0.604	0.623	0.634	0.650	0.662	0.671	0.683	0.692
10	0.558	0.585	0.602	0.621	0.636	0.649	0.660	0.673	0.684	0.692

Table 42: Accuracy with judge and expert competences from exponential distribution with scale 10 truncated over (0.5, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.551	0.551	0.577	0.577	0.595	0.596	0.610	0.607	0.628	0.621
1	0.551	0.574	0.597	0.610	0.621	0.637	0.645	0.656	0.663	0.671
2	0.549	0.577	0.596	0.609	0.625	0.637	0.645	0.657	0.663	0.674
3	0.551	0.576	0.594	0.610	0.624	0.635	0.646	0.654	0.666	0.675
4	0.552	0.578	0.595	0.609	0.624	0.635	0.645	0.656	0.667	0.675
5	0.548	0.574	0.595	0.608	0.620	0.636	0.648	0.657	0.663	0.674
6	0.549	0.576	0.593	0.610	0.621	0.638	0.648	0.653	0.666	0.676
7	0.551	0.576	0.593	0.612	0.620	0.634	0.647	0.657	0.663	0.673
8	0.552	0.576	0.594	0.611	0.622	0.635	0.645	0.656	0.667	0.674
9	0.552	0.575	0.593	0.608	0.623	0.638	0.644	0.652	0.666	0.672
10	0.552	0.576	0.595	0.610	0.624	0.635	0.648	0.655	0.664	0.674

Table 43: Accuracy with judge and expert competences from exponential distribution with scale 1 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.418	0.415	0.376	0.377	0.349	0.349	0.322	0.326	0.302	0.307
1	0.437	0.418	0.406	0.397	0.392	0.390	0.385	0.384	0.381	0.380
2	0.420	0.391	0.376	0.364	0.357	0.352	0.346	0.345	0.345	0.340
3	0.401	0.369	0.349	0.338	0.330	0.323	0.319	0.314	0.315	0.308
4	0.388	0.353	0.327	0.314	0.306	0.297	0.294	0.291	0.290	0.285
5	0.375	0.337	0.315	0.297	0.288	0.280	0.274	0.269	0.267	0.263
6	0.367	0.323	0.296	0.283	0.271	0.261	0.255	0.251	0.247	0.244
7	0.358	0.312	0.285	0.268	0.253	0.244	0.240	0.235	0.232	0.229
8	0.349	0.300	0.273	0.253	0.243	0.230	0.224	0.221	0.217	0.215
9	0.343	0.289	0.263	0.244	0.228	0.221	0.214	0.207	0.201	0.200
10	0.336	0.283	0.251	0.233	0.221	0.209	0.200	0.195	0.193	0.188

Table 44: Accuracy with judge and expert competences from exponential distribution with scale 2 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.342	0.345	0.274	0.272	0.226	0.225	0.188	0.188	0.162	0.158
1	0.375	0.336	0.315	0.297	0.292	0.285	0.282	0.276	0.274	0.274
2	0.341	0.294	0.261	0.245	0.235	0.225	0.219	0.217	0.212	0.210
3	0.316	0.261	0.227	0.207	0.194	0.183	0.177	0.170	0.167	0.165
4	0.299	0.234	0.202	0.176	0.162	0.152	0.144	0.138	0.134	0.131
5	0.283	0.218	0.181	0.155	0.139	0.128	0.120	0.116	0.111	0.108
6	0.271	0.204	0.165	0.139	0.124	0.110	0.101	0.096	0.090	0.089
7	0.266	0.192	0.152	0.126	0.107	0.096	0.085	0.080	0.077	0.074
8	0.261	0.184	0.143	0.114	0.098	0.084	0.077	0.070	0.065	0.063
9	0.255	0.178	0.133	0.107	0.090	0.074	0.066	0.060	0.055	0.051
10	0.250	0.173	0.129	0.099	0.082	0.068	0.058	0.053	0.049	0.045

Table 45: Accuracy with judge and expert competences from exponential distribution with scale 3 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.283	0.281	0.190	0.190	0.138	0.139	0.103	0.102	0.076	0.078
1	0.316	0.264	0.235	0.219	0.206	0.197	0.192	0.188	0.187	0.187
2	0.278	0.214	0.177	0.156	0.142	0.133	0.127	0.124	0.117	0.116
3	0.253	0.181	0.143	0.119	0.104	0.093	0.088	0.082	0.079	0.077
4	0.241	0.164	0.123	0.098	0.079	0.068	0.062	0.057	0.053	0.052
5	0.229	0.150	0.108	0.083	0.066	0.053	0.047	0.041	0.039	0.035
6	0.221	0.144	0.100	0.074	0.056	0.045	0.038	0.031	0.028	0.025
7	0.219	0.139	0.097	0.067	0.050	0.039	0.031	0.025	0.021	0.019
8	0.218	0.135	0.091	0.063	0.046	0.034	0.025	0.022	0.018	0.015
9	0.215	0.134	0.088	0.062	0.043	0.030	0.024	0.018	0.015	0.012
10	0.215	0.132	0.088	0.060	0.041	0.029	0.022	0.016	0.013	0.010

Table 46: Accuracy with judge and expert competences from exponential distribution with scale 4 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.230	0.232	0.136	0.137	0.085	0.084	0.054	0.055	0.037	0.035
1	0.264	0.206	0.174	0.153	0.142	0.134	0.128	0.123	0.122	0.121
2	0.226	0.156	0.118	0.096	0.082	0.073	0.067	0.063	0.062	0.061
3	0.208	0.133	0.093	0.069	0.055	0.045	0.040	0.035	0.033	0.030
4	0.199	0.120	0.081	0.056	0.041	0.032	0.026	0.021	0.019	0.017
5	0.196	0.115	0.074	0.050	0.034	0.025	0.019	0.015	0.013	0.010
6	0.193	0.112	0.073	0.046	0.031	0.022	0.015	0.011	0.009	0.007
7	0.191	0.113	0.071	0.046	0.029	0.020	0.013	0.010	0.007	0.005
8	0.192	0.110	0.070	0.044	0.029	0.018	0.014	0.009	0.006	0.005
9	0.191	0.111	0.068	0.044	0.029	0.018	0.012	0.008	0.006	0.004
10	0.192	0.111	0.069	0.044	0.028	0.019	0.013	0.008	0.006	0.004

Table 47: Accuracy with judge and expert competences from exponential distribution with scale 5 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.192	0.194	0.097	0.096	0.052	0.053	0.030	0.030	0.017	0.016
1	0.217	0.158	0.125	0.103	0.093	0.086	0.083	0.081	0.078	0.078
2	0.187	0.118	0.080	0.059	0.047	0.038	0.035	0.032	0.029	0.029
3	0.177	0.103	0.065	0.042	0.031	0.022	0.017	0.015	0.013	0.012
4	0.169	0.097	0.060	0.036	0.024	0.016	0.012	0.009	0.007	0.006
5	0.169	0.096	0.057	0.035	0.021	0.013	0.009	0.006	0.004	0.003
6	0.171	0.092	0.057	0.034	0.021	0.012	0.008	0.005	0.003	0.002
7	0.170	0.096	0.055	0.034	0.020	0.013	0.008	0.005	0.003	0.002
8	0.168	0.095	0.055	0.033	0.020	0.012	0.008	0.006	0.003	0.002
9	0.169	0.095	0.056	0.033	0.020	0.013	0.008	0.005	0.003	0.002
10	0.170	0.095	0.055	0.033	0.021	0.013	0.008	0.005	0.003	0.002

Table 48: Accuracy with judge and expert competences from exponential distribution with scale 6 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.163	0.165	0.071	0.072	0.034	0.034	0.017	0.017	0.008	0.008
1	0.184	0.121	0.089	0.070	0.062	0.056	0.051	0.051	0.048	0.049
2	0.159	0.091	0.057	0.039	0.028	0.021	0.018	0.015	0.014	0.013
3	0.153	0.085	0.049	0.028	0.018	0.012	0.008	0.006	0.005	0.004
4	0.151	0.082	0.046	0.026	0.016	0.010	0.006	0.004	0.003	0.002
5	0.150	0.081	0.046	0.025	0.015	0.009	0.005	0.003	0.002	0.001
6	0.153	0.083	0.044	0.025	0.015	0.009	0.005	0.003	0.002	0.001
7	0.151	0.081	0.046	0.025	0.015	0.009	0.005	0.003	0.002	0.001
8	0.150	0.080	0.046	0.026	0.015	0.009	0.005	0.003	0.002	0.001
9	0.152	0.082	0.045	0.026	0.015	0.009	0.005	0.003	0.002	0.001
10	0.150	0.083	0.045	0.026	0.015	0.008	0.005	0.003	0.002	0.001

Table 49: Accuracy with judge and expert competences from exponential distribution with scale 7 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.143	0.142	0.054	0.056	0.022	0.023	0.010	0.010	0.004	0.004
1	0.155	0.095	0.065	0.048	0.040	0.035	0.034	0.031	0.030	0.030
2	0.138	0.077	0.043	0.025	0.016	0.011	0.009	0.008	0.007	0.006
3	0.136	0.071	0.038	0.021	0.012	0.007	0.004	0.003	0.002	0.002
4	0.137	0.071	0.037	0.020	0.011	0.006	0.004	0.002	0.001	0.001
5	0.136	0.072	0.038	0.019	0.011	0.006	0.003	0.002	0.001	0.001
6	0.134	0.071	0.037	0.019	0.011	0.006	0.004	0.002	0.001	0.001
7	0.135	0.070	0.037	0.020	0.011	0.006	0.004	0.002	0.001	0.001
8	0.132	0.070	0.037	0.020	0.011	0.006	0.003	0.002	0.001	0.001
9	0.134	0.070	0.038	0.020	0.011	0.006	0.004	0.002	0.001	0.001
10	0.135	0.072	0.038	0.020	0.011	0.006	0.003	0.002	0.001	0.001

Table 50: Accuracy with judge and expert competences from exponential distribution with scale 8 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.126	0.125	0.043	0.042	0.016	0.015	0.006	0.006	0.003	0.003
1	0.135	0.079	0.048	0.033	0.026	0.022	0.021	0.020	0.019	0.018
2	0.122	0.064	0.034	0.018	0.011	0.007	0.005	0.004	0.003	0.003
3	0.120	0.062	0.031	0.016	0.009	0.005	0.002	0.002	0.001	0.001
4	0.119	0.062	0.032	0.016	0.008	0.004	0.002	0.001	0.001	0.000
5	0.119	0.062	0.031	0.016	0.009	0.004	0.002	0.001	0.001	0.001
6	0.121	0.061	0.030	0.015	0.008	0.004	0.003	0.001	0.001	0.000
7	0.120	0.063	0.032	0.017	0.009	0.004	0.002	0.001	0.001	0.000
8	0.121	0.062	0.032	0.015	0.009	0.004	0.003	0.001	0.001	0.000
9	0.122	0.062	0.033	0.015	0.008	0.004	0.002	0.001	0.001	0.000
10	0.120	0.063	0.031	0.016	0.008	0.004	0.002	0.001	0.001	0.000

Table 51: Accuracy with judge and expert competences from exponential distribution with scale 9 truncated over (0.0, 1.0)

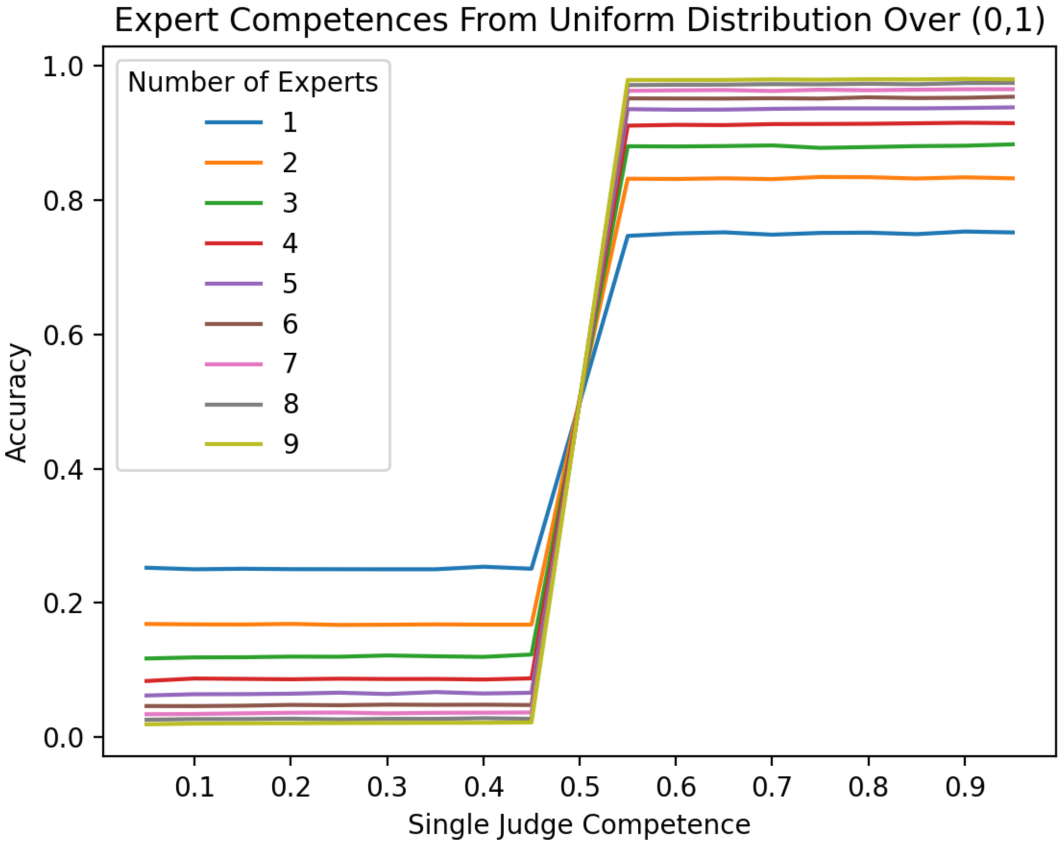
m	1	2	3	4	5	6	7	8	9	10
n										
0	0.108	0.112	0.034	0.034	0.011	0.011	0.004	0.004	0.001	0.001
1	0.117	0.065	0.038	0.023	0.017	0.014	0.013	0.011	0.011	0.011
2	0.110	0.057	0.026	0.014	0.007	0.004	0.003	0.002	0.002	0.001
3	0.109	0.055	0.026	0.013	0.007	0.003	0.002	0.001	0.001	0.000
4	0.109	0.055	0.026	0.012	0.006	0.003	0.002	0.001	0.001	0.000
5	0.109	0.055	0.026	0.012	0.006	0.003	0.001	0.001	0.000	0.000
6	0.107	0.054	0.027	0.013	0.006	0.003	0.001	0.001	0.000	0.000
7	0.109	0.055	0.026	0.012	0.006	0.003	0.002	0.001	0.000	0.000
8	0.110	0.055	0.026	0.013	0.006	0.003	0.002	0.001	0.001	0.000
9	0.110	0.055	0.026	0.012	0.006	0.003	0.002	0.001	0.000	0.000
10	0.109	0.055	0.026	0.013	0.006	0.003	0.002	0.001	0.000	0.000

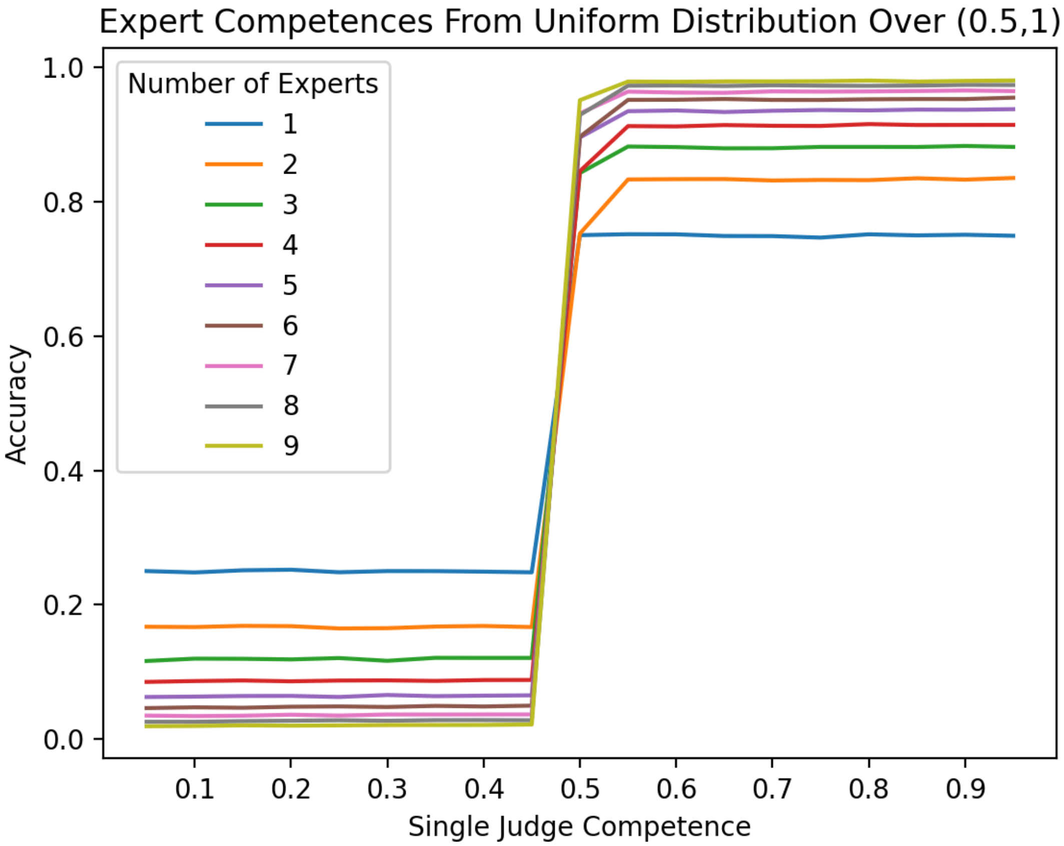
Table 52: Accuracy with judge and expert competences from exponential distribution with scale 10 truncated over (0.0, 1.0)

m	1	2	3	4	5	6	7	8	9	10
n										
0	0.101	0.099	0.028	0.027	0.009	0.009	0.003	0.003	0.001	0.001
1	0.104	0.056	0.029	0.017	0.012	0.009	0.008	0.007	0.007	0.007
2	0.098	0.050	0.022	0.011	0.005	0.003	0.001	0.001	0.001	0.001
3	0.098	0.049	0.022	0.010	0.005	0.002	0.001	0.001	0.000	0.000
4	0.097	0.050	0.023	0.010	0.005	0.002	0.001	0.001	0.000	0.000
5	0.098	0.049	0.023	0.010	0.004	0.002	0.001	0.000	0.000	0.000
6	0.100	0.049	0.023	0.010	0.005	0.003	0.001	0.000	0.000	0.000
7	0.097	0.051	0.022	0.010	0.005	0.002	0.001	0.001	0.000	0.000
8	0.098	0.050	0.022	0.010	0.005	0.002	0.001	0.001	0.000	0.000
9	0.098	0.050	0.022	0.010	0.005	0.002	0.001	0.001	0.000	0.000
10	0.098	0.050	0.022	0.010	0.005	0.002	0.001	0.001	0.000	0.000

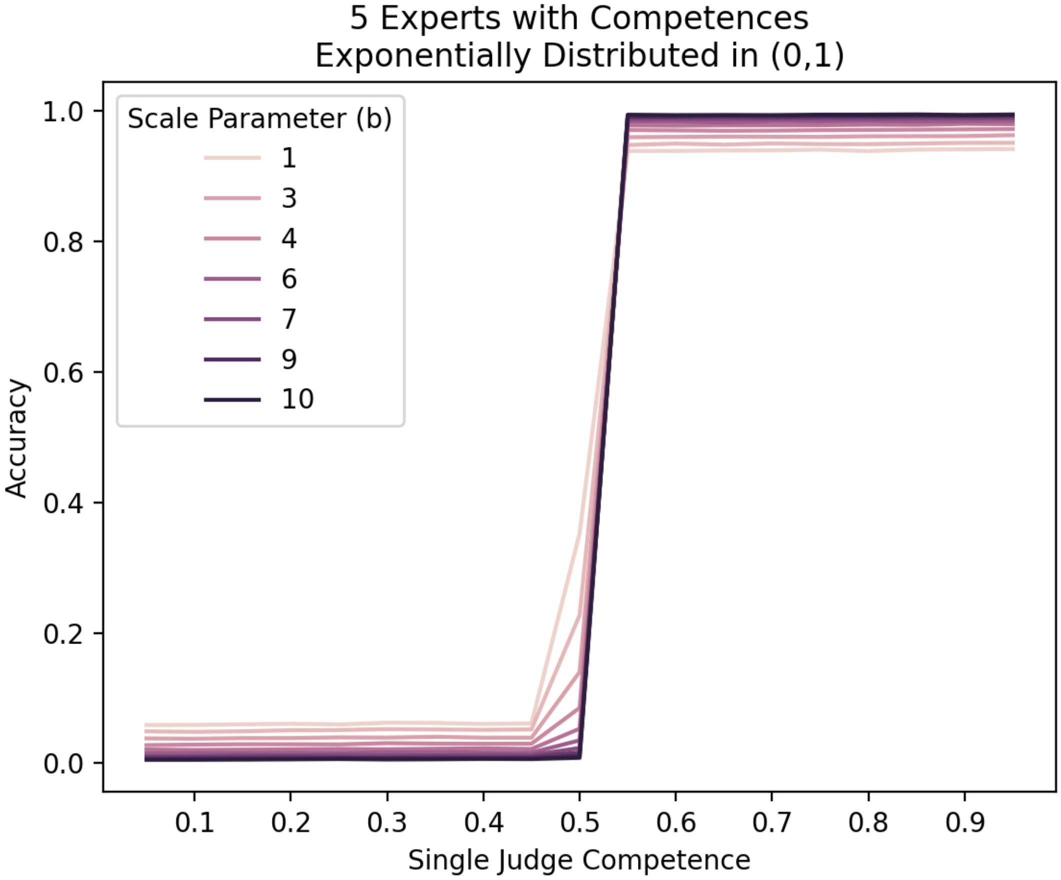
A SINGLE JUDGE

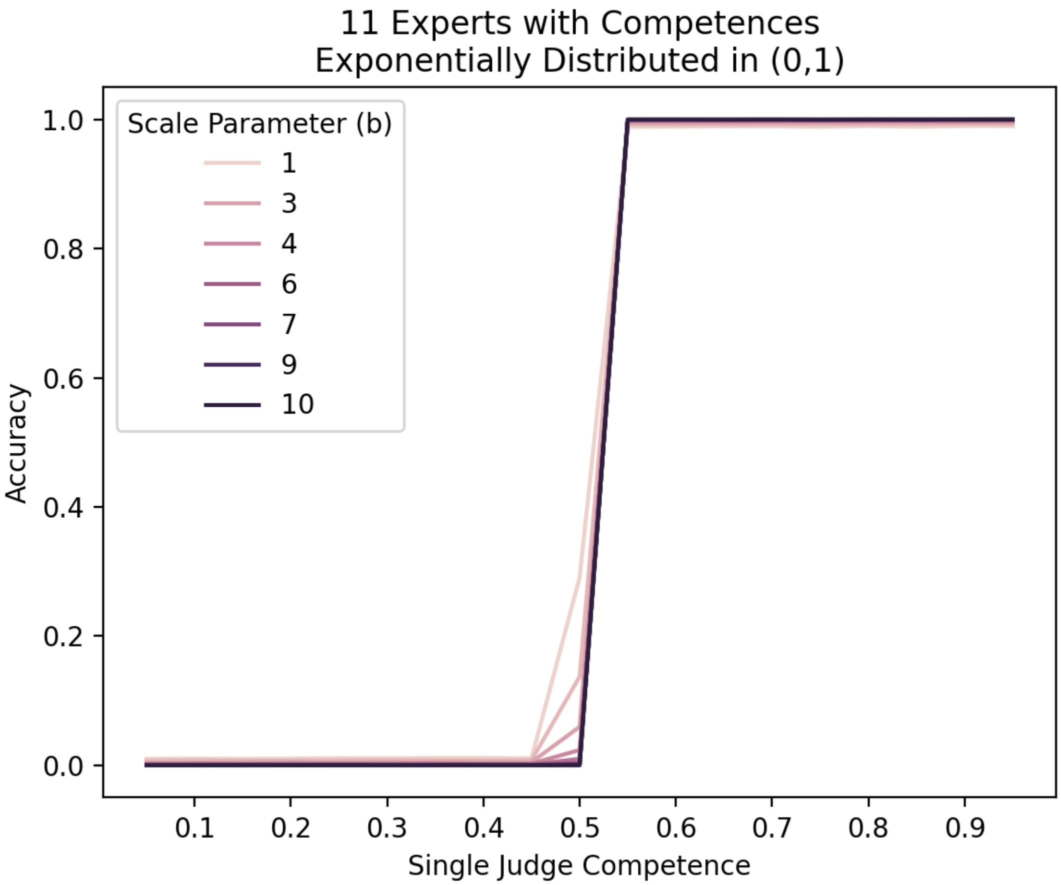
A.1 Uniformly Distributed competences

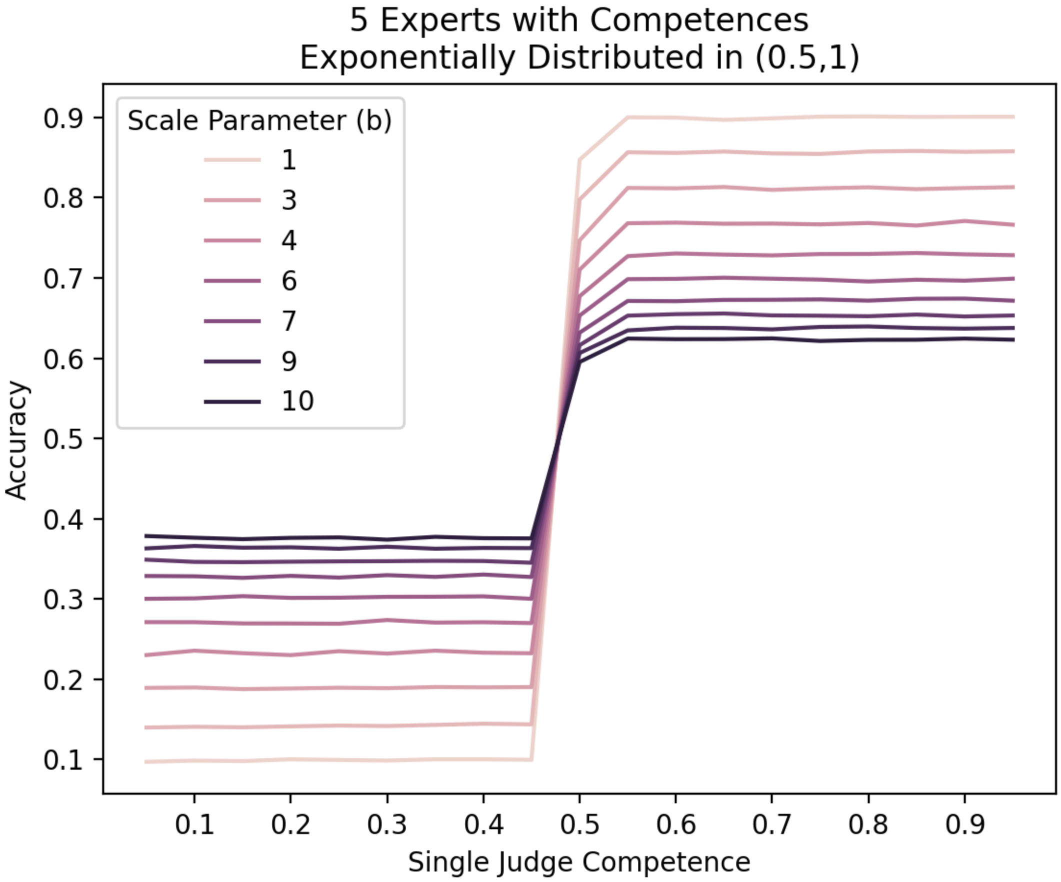


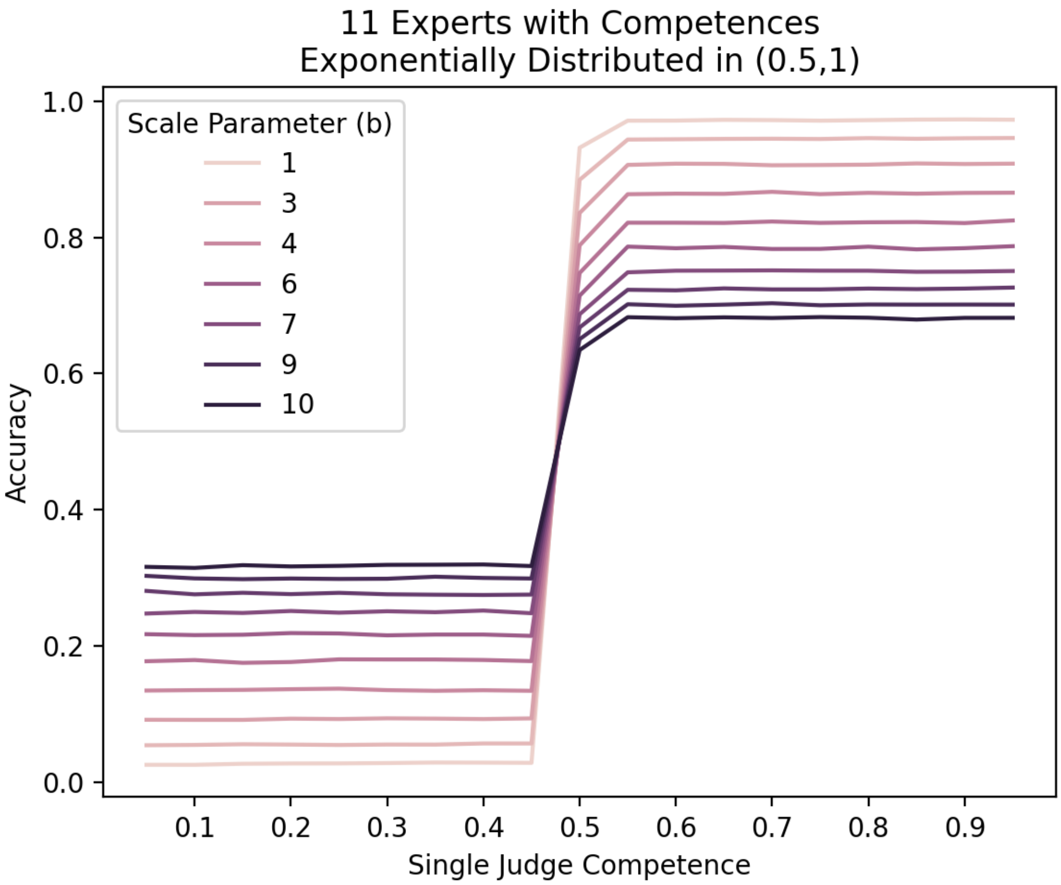

Expert Competences From Uniform Distribution Over (0.5,1)
Number of Experts
1
2
3
4
5
6
7
8
9
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

A.2 Exponentially Distributed competences

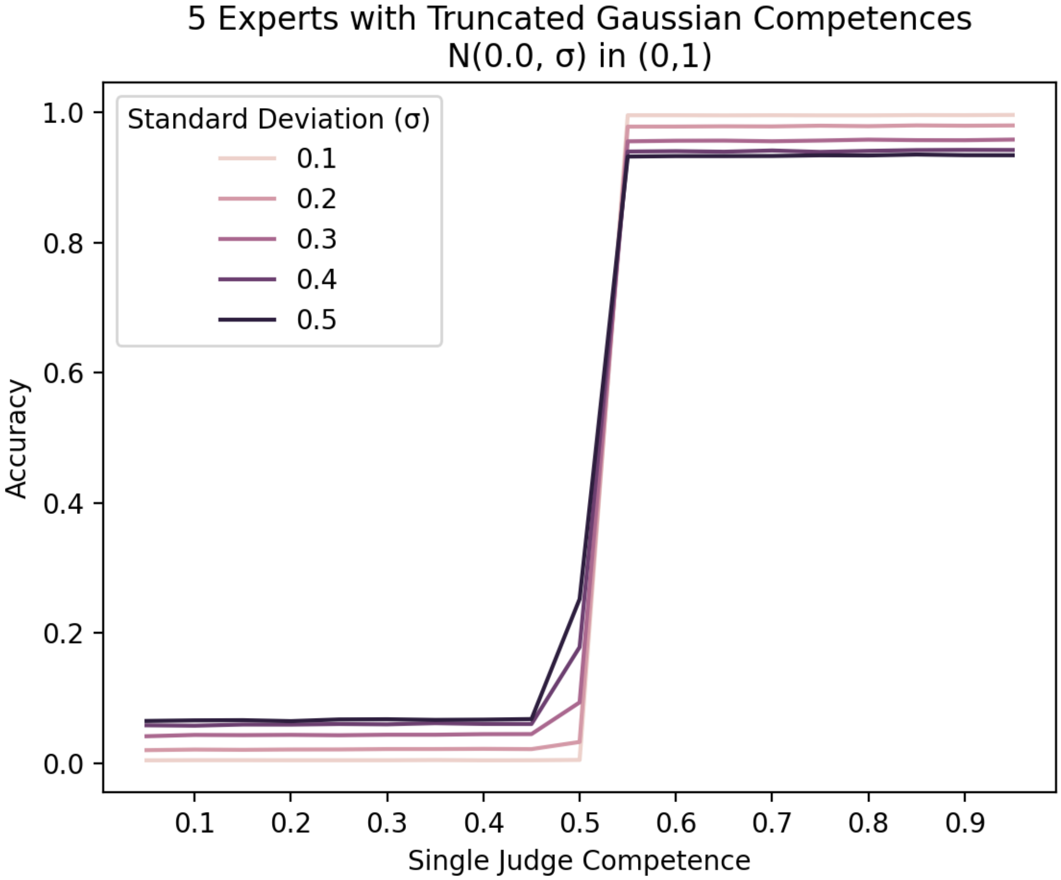


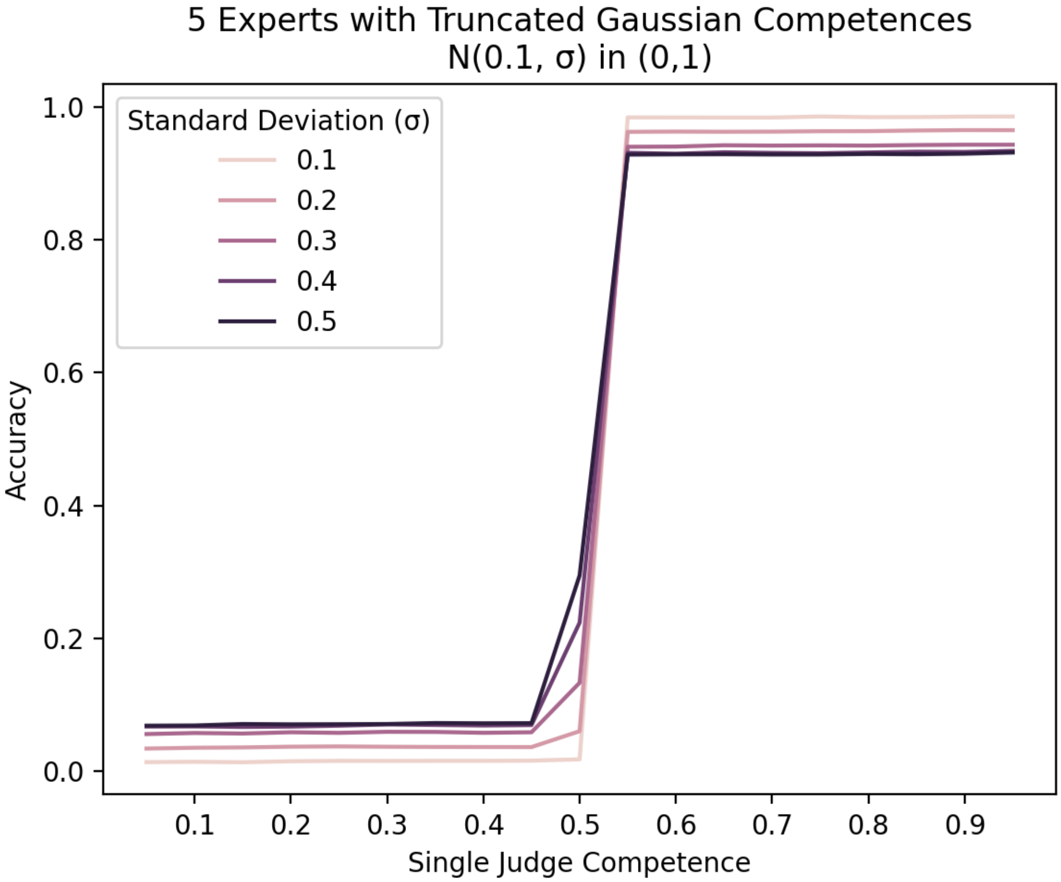

11 Experts with Competences
Exponentially Distributed in (0,1)
Scale Parameter (b)
1
3
4
6
7
9
10
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

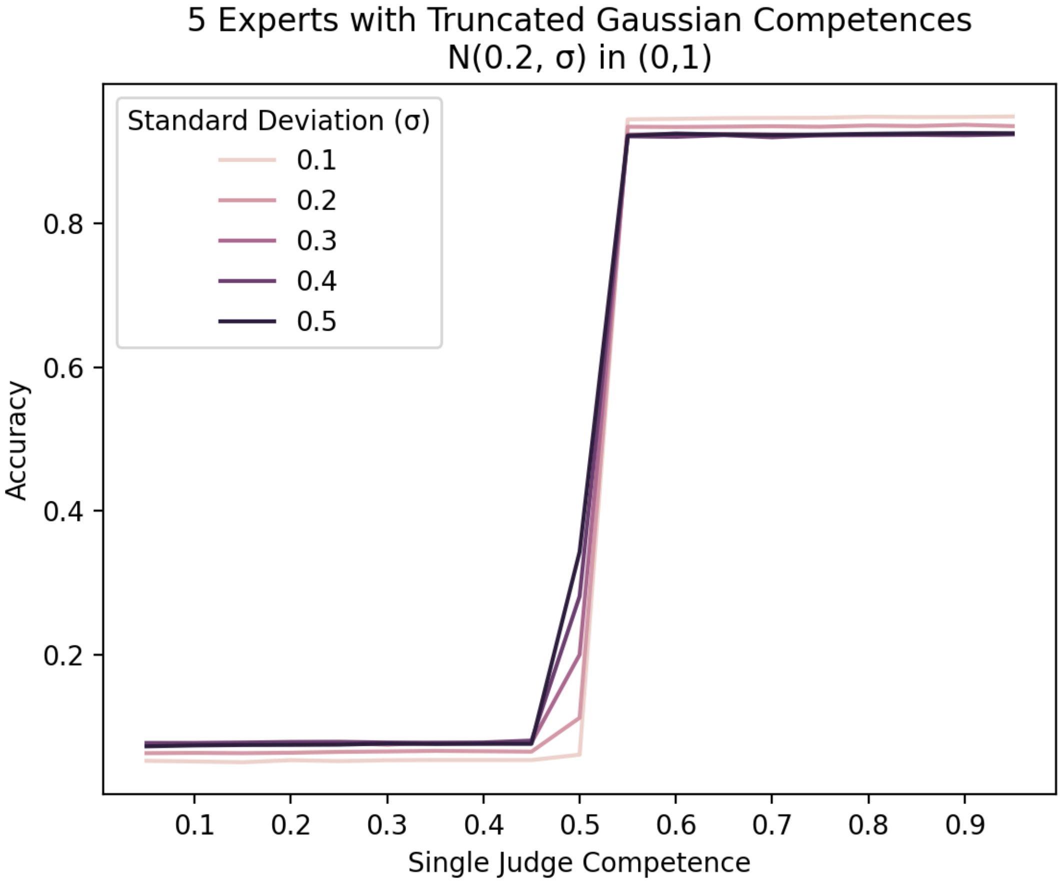

5 Experts with Competences
Exponentially Distributed in (0.5,1)
Scale Parameter (b)
1
3
4
6
7
9
10
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

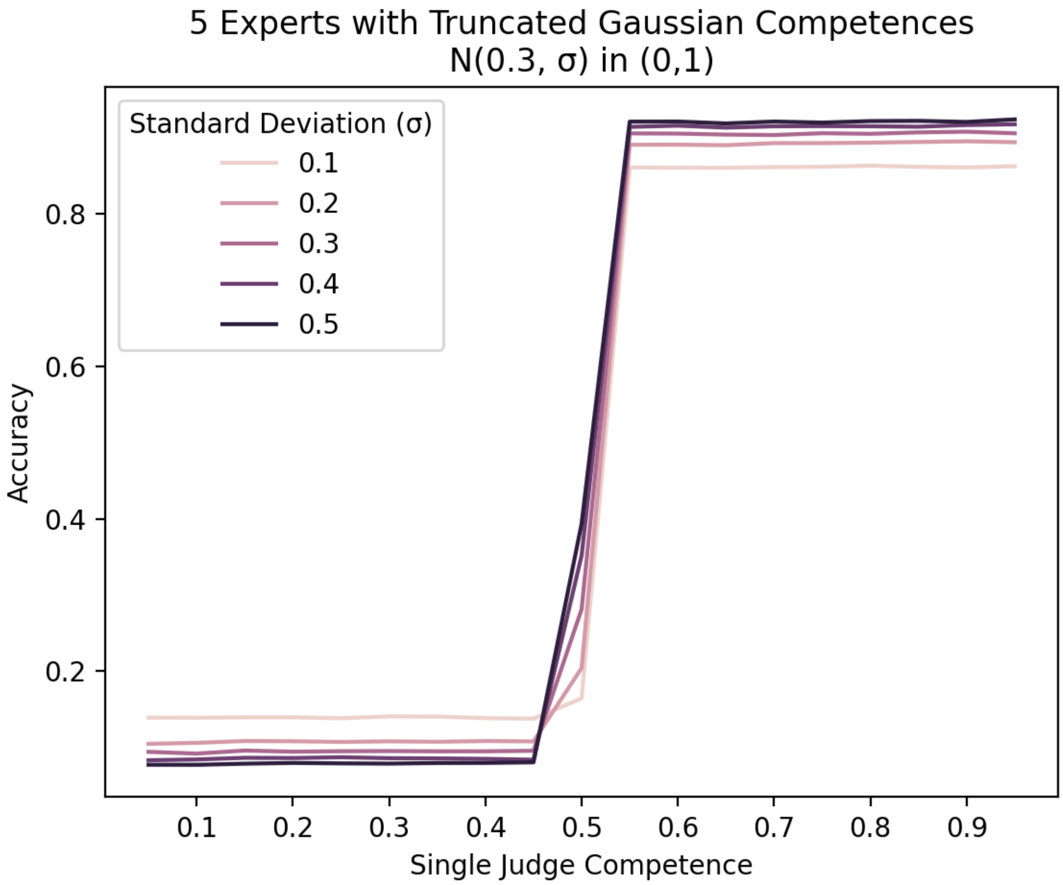

11 Experts with Competences
Exponentially Distributed in (0.5,1)
Scale Parameter (b)
1
3
4
6
7
9
10
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

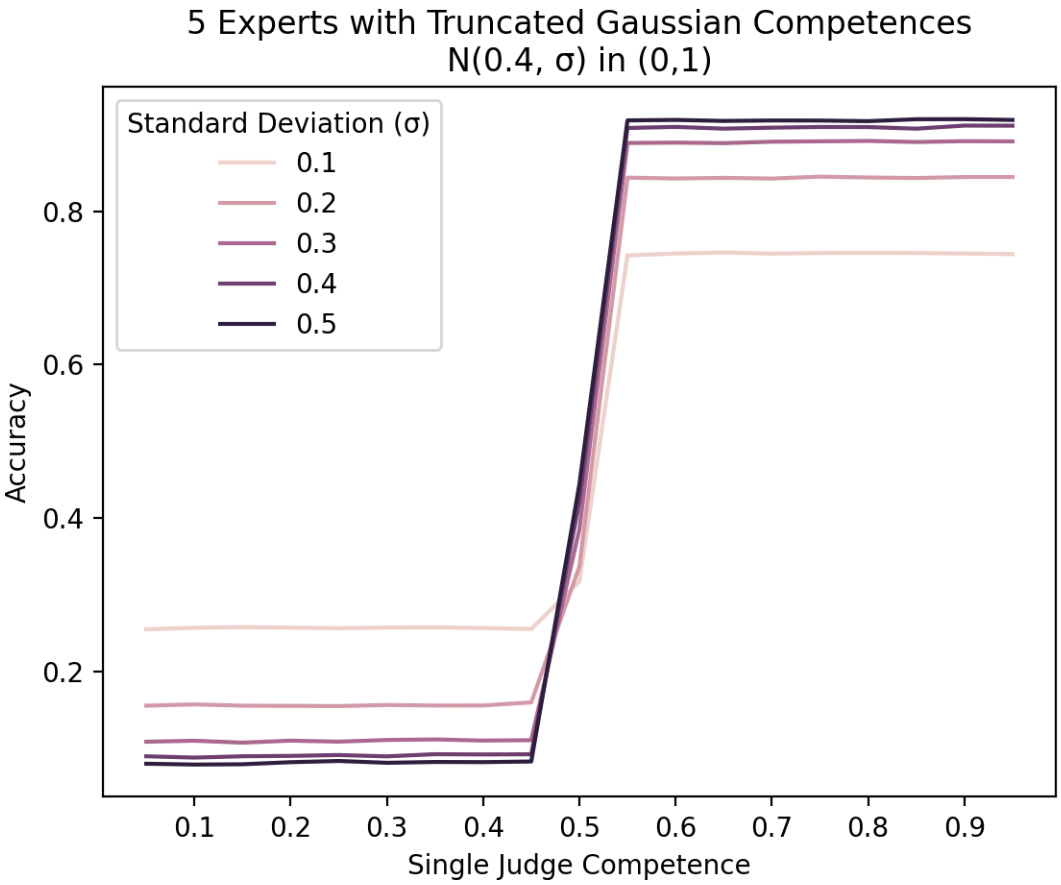
A.3 Normally Distributed competences in $(0, 1)$ with 5 Experts

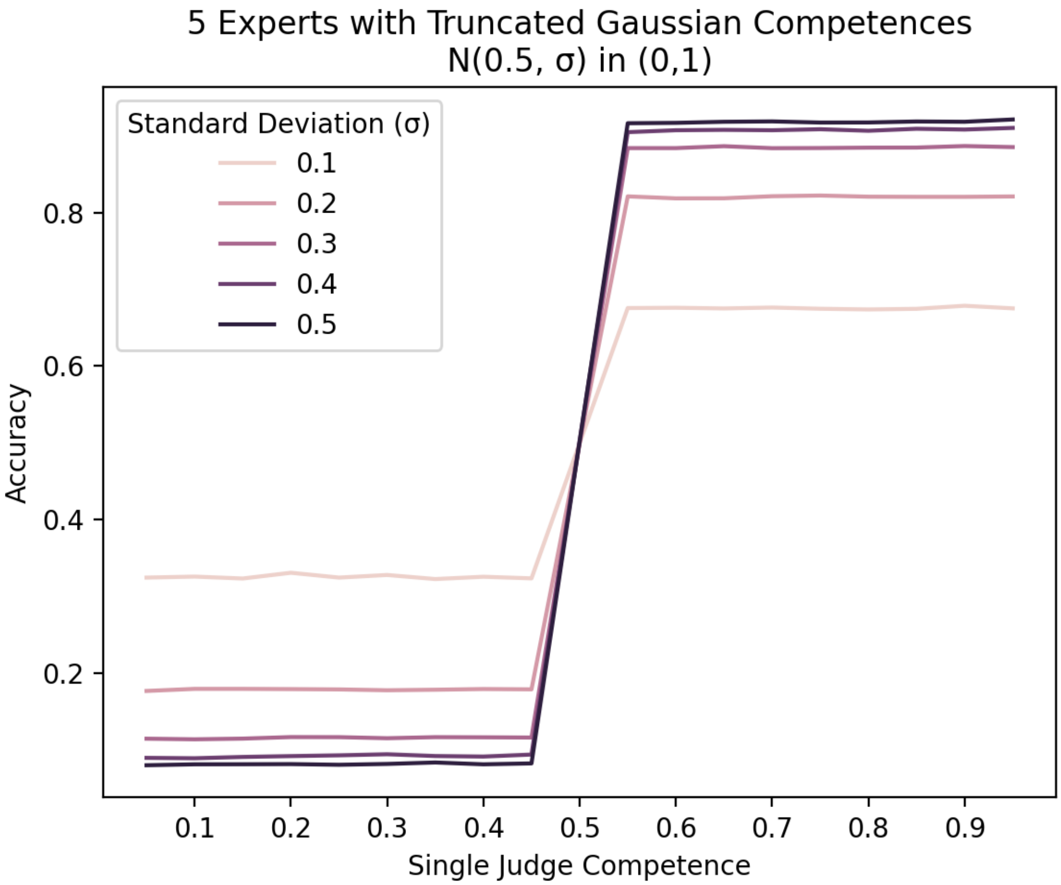


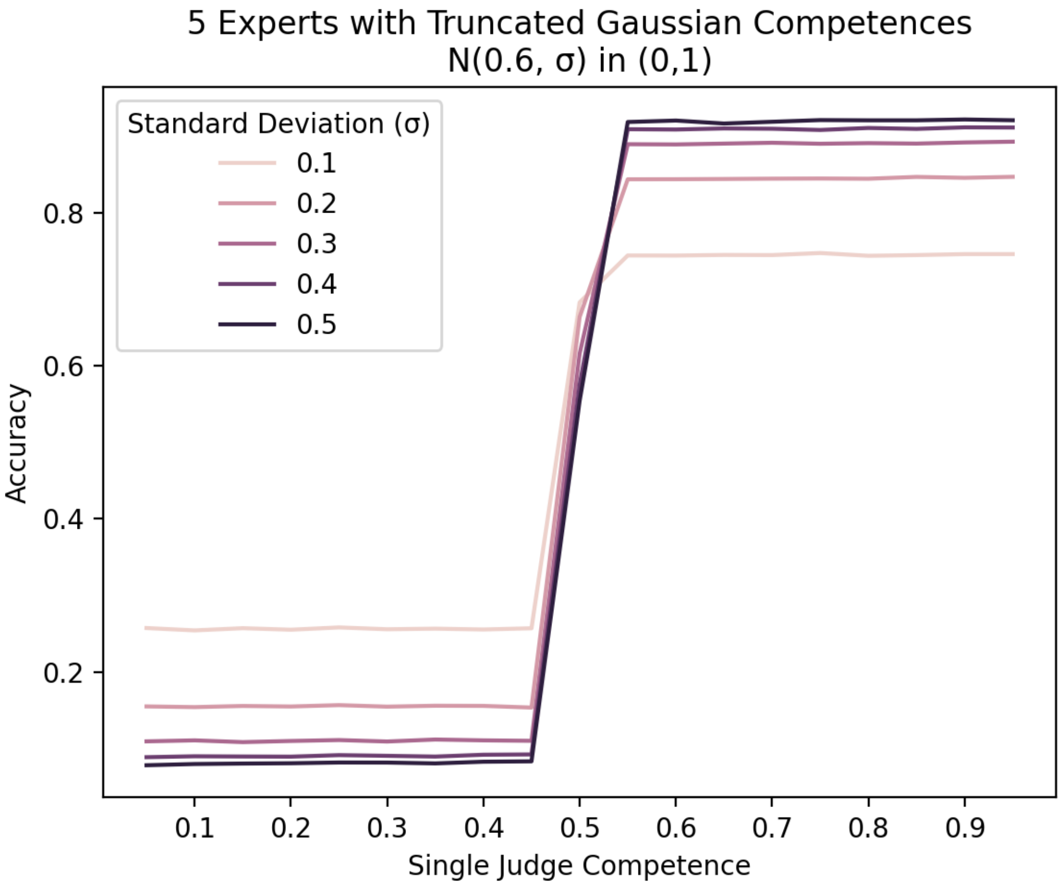

5 Experts with Truncated Gaussian Competences
N(0.1, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

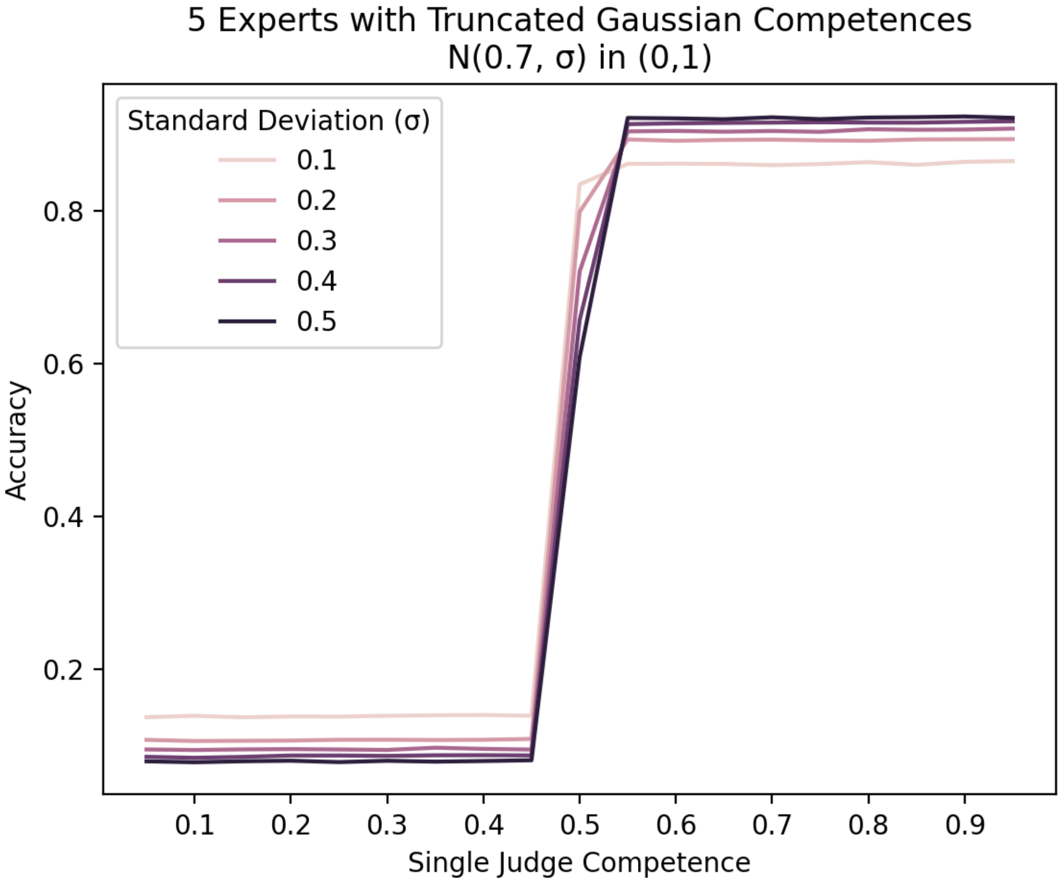

5 Experts with Truncated Gaussian Competences
N(0.2, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

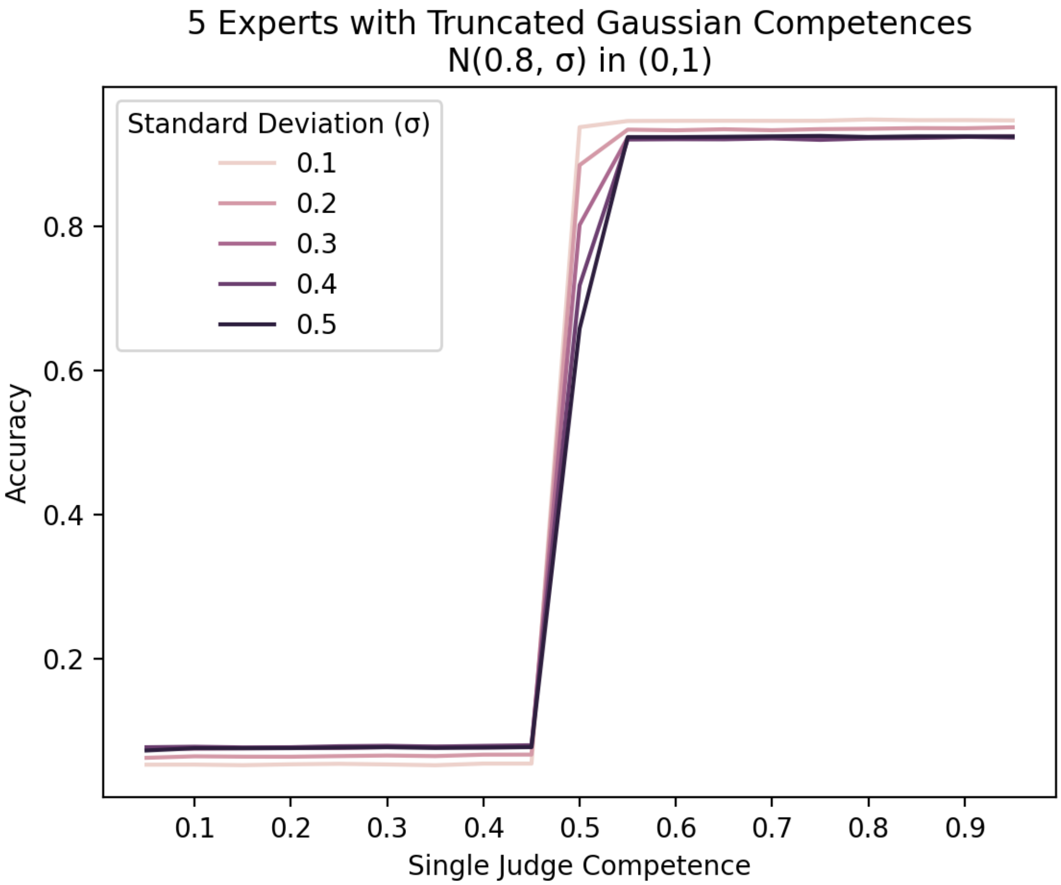

5 Experts with Truncated Gaussian Competences
N(0.3, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

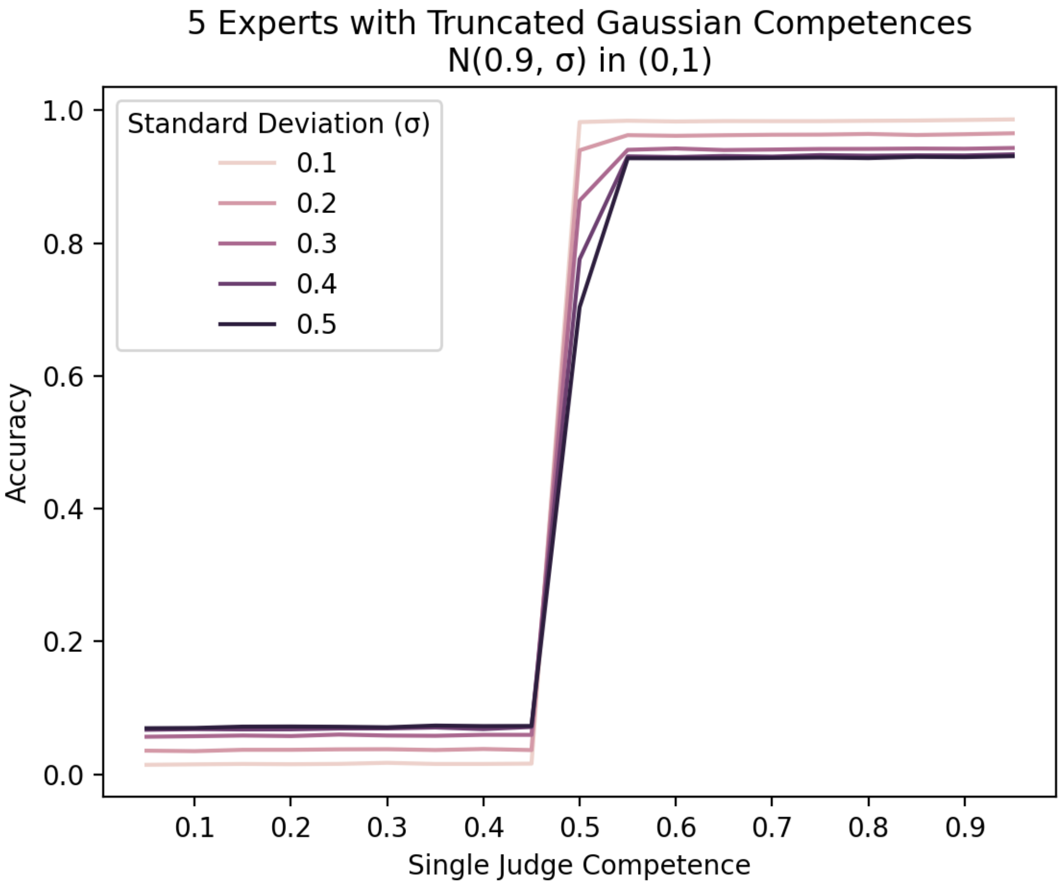

5 Experts with Truncated Gaussian Competences
N(0.4, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

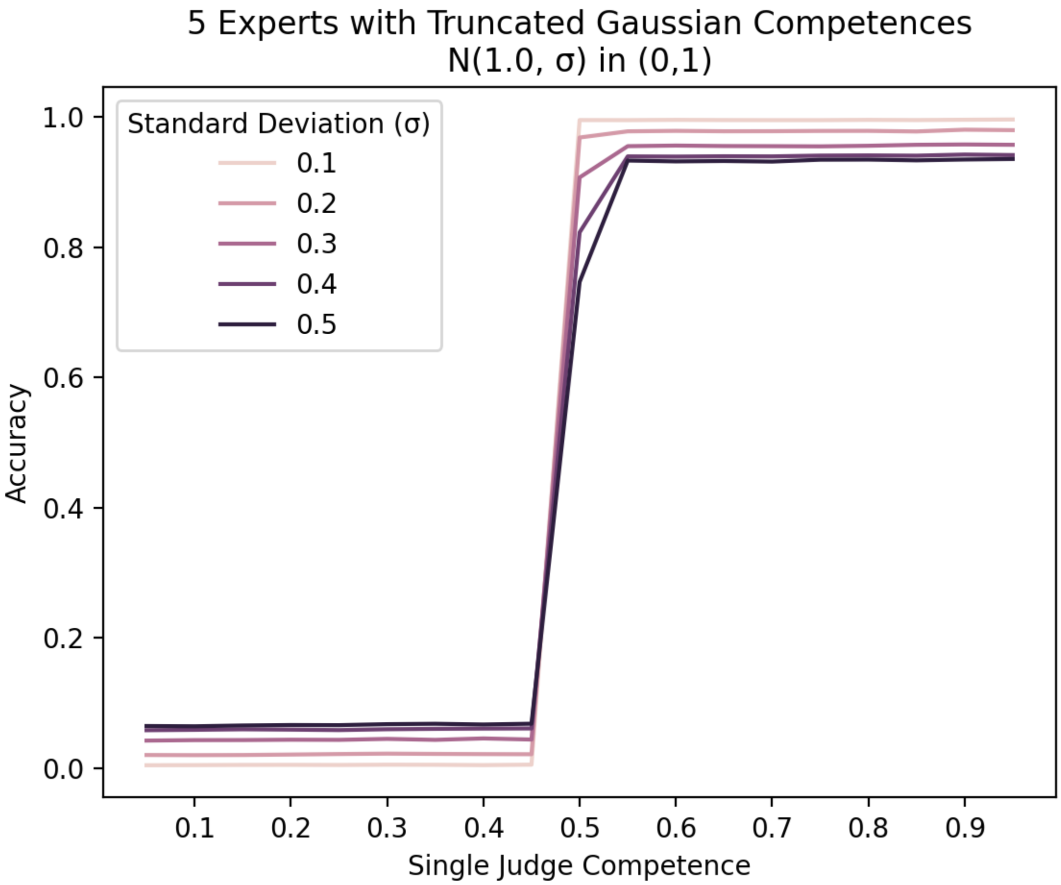

5 Experts with Truncated Gaussian Competences
N(0.5, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

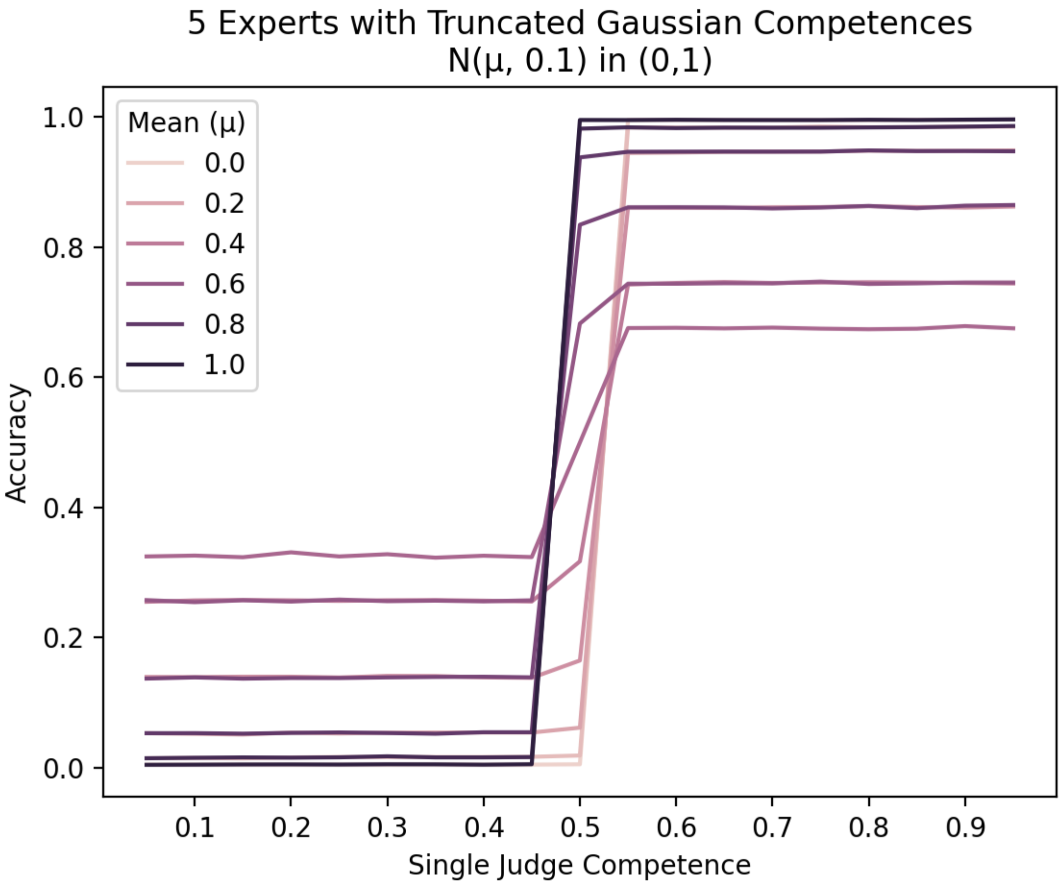

5 Experts with Truncated Gaussian Competences
N(0.6, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

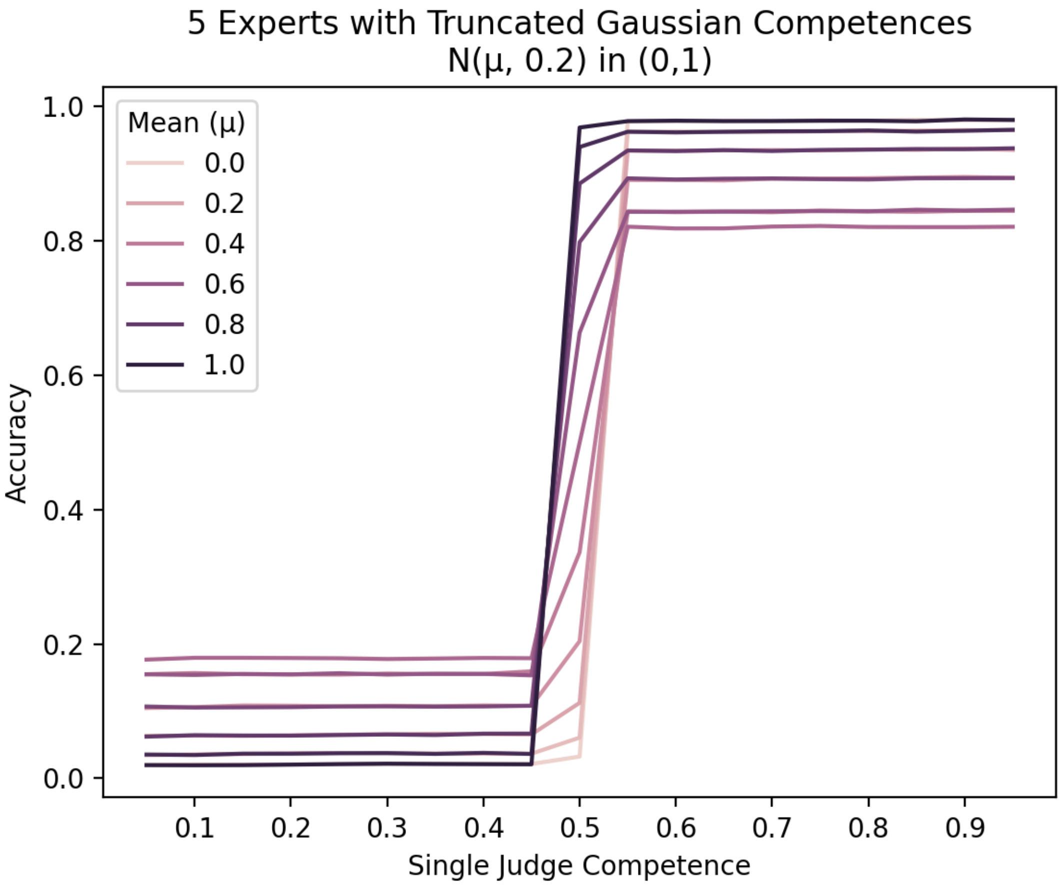

5 Experts with Truncated Gaussian Competences
N(0.7, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.8
0.6
0.4
0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

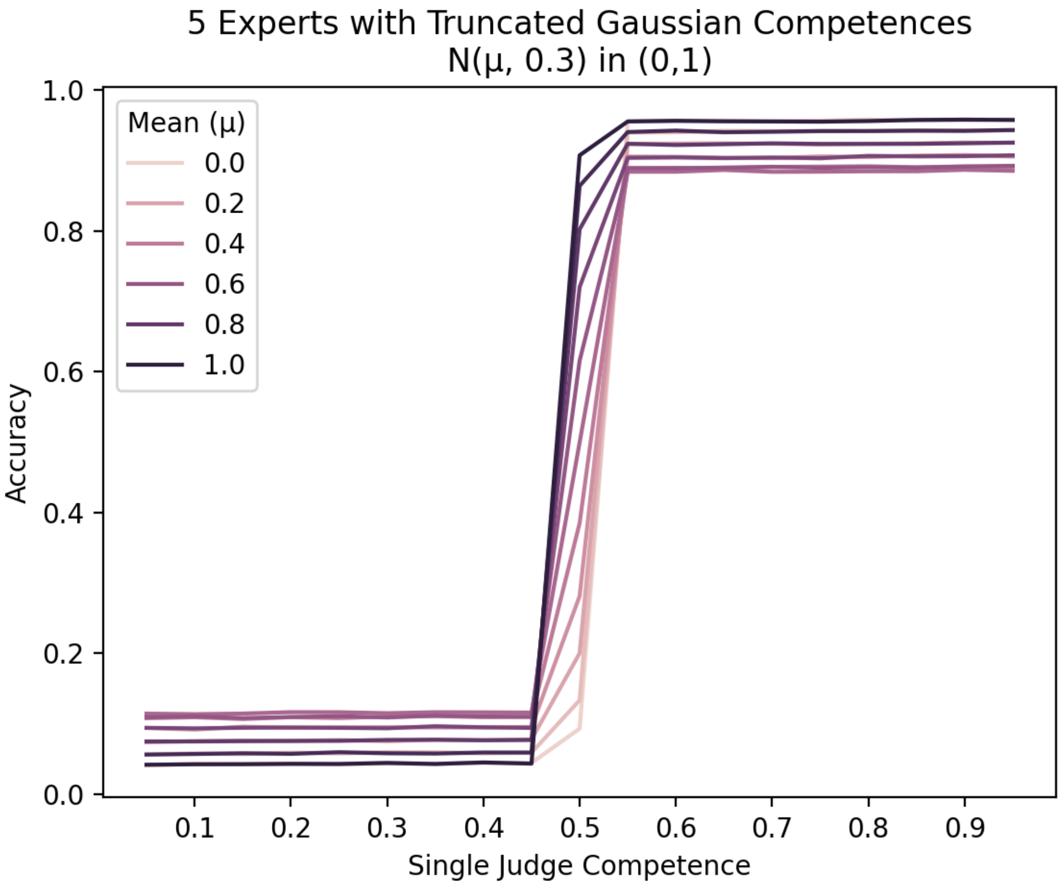

5 Experts with Truncated Gaussian Competences
N(0.8, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

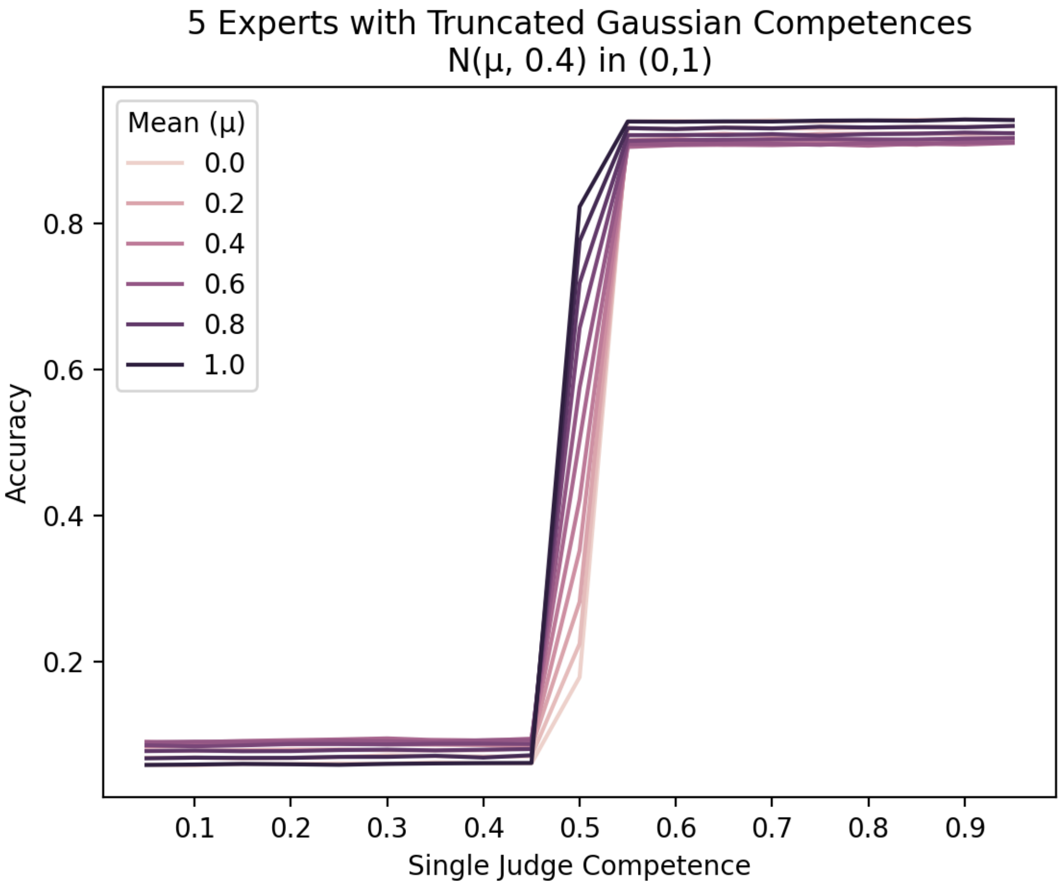

5 Experts with Truncated Gaussian Competences
N(0.9, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

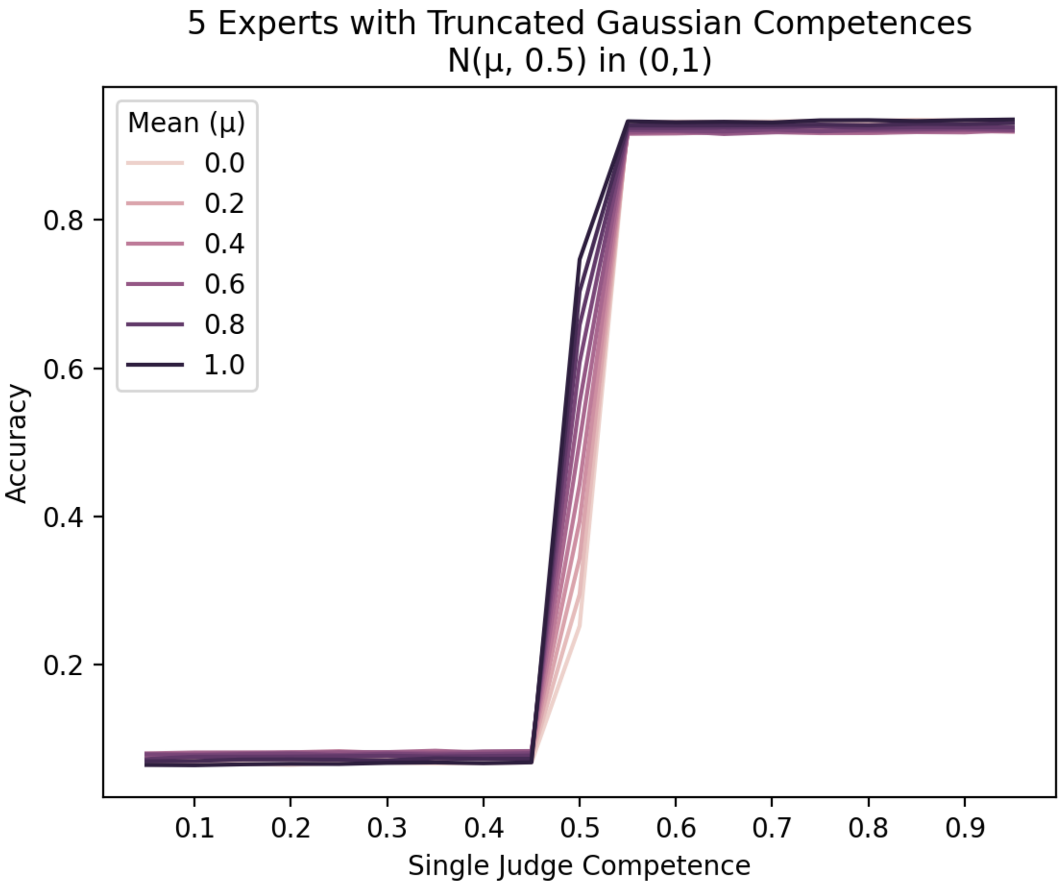
5 Experts with Truncated Gaussian Competences
N(1.0, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence


5 Experts with Truncated Gaussian Competences
N(μ, 0.1) in (0,1)
Mean (μ)
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

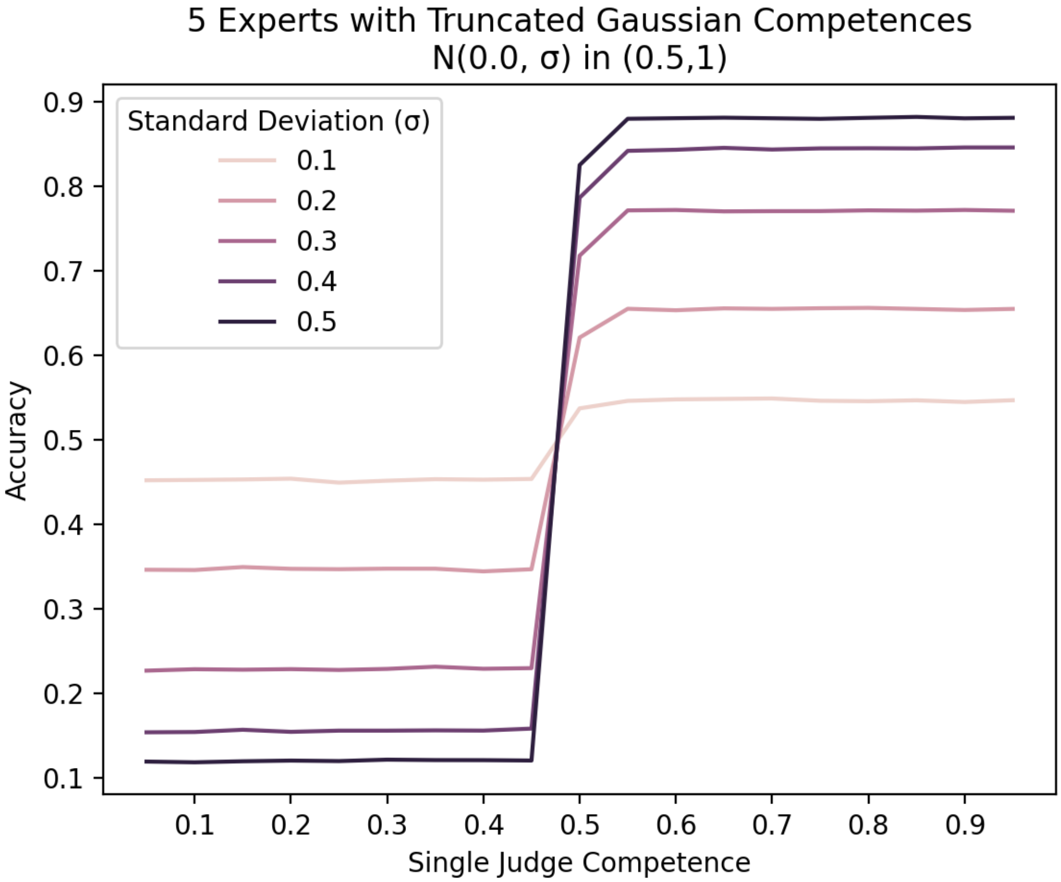

5 Experts with Truncated Gaussian Competences
N(μ, 0.2) in (0,1)
Mean (μ)
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

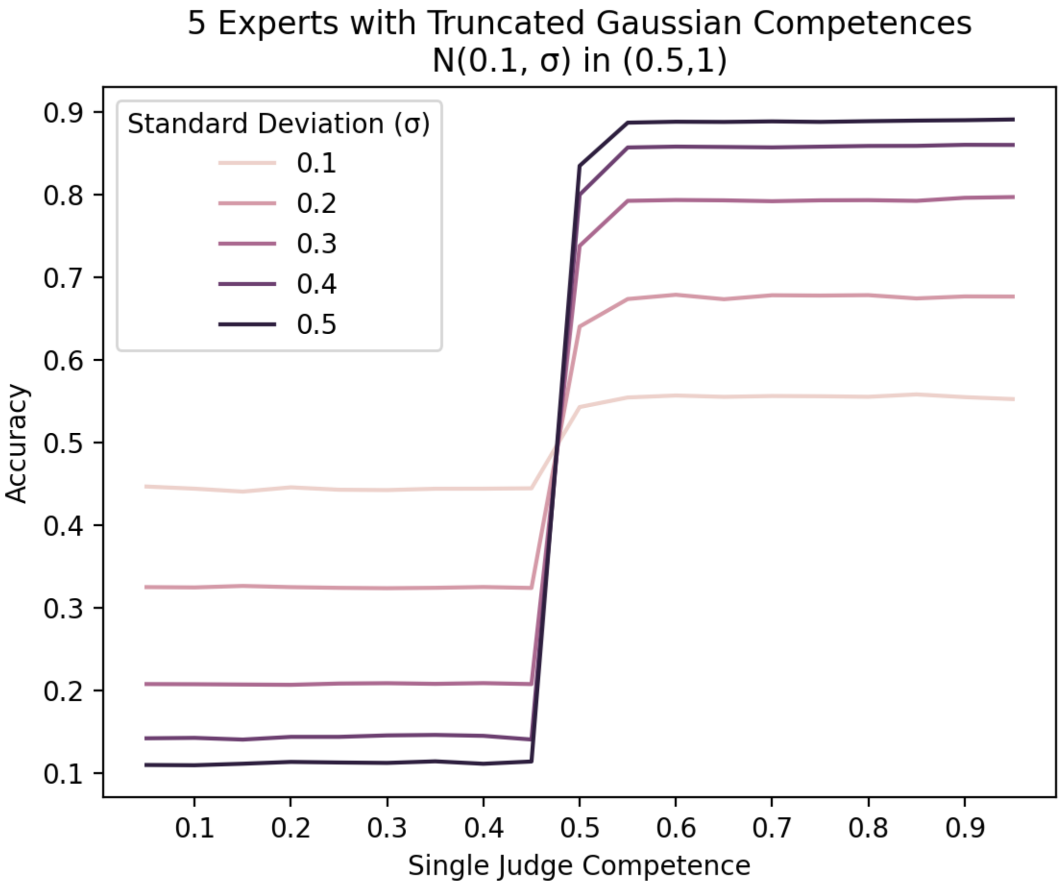

5 Experts with Truncated Gaussian Competences
N(μ, 0.3) in (0,1)
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Mean (μ)
0.0
0.2
0.4
0.6
0.8
1.0

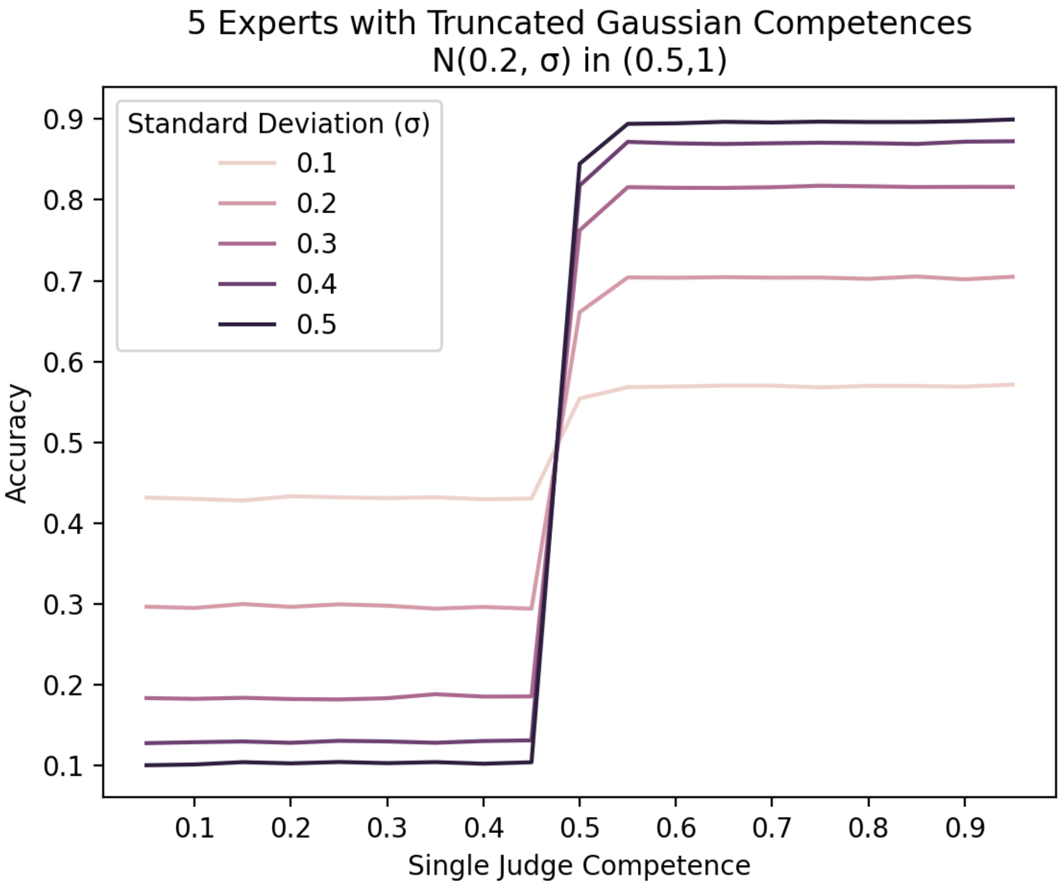

5 Experts with Truncated Gaussian Competences
N(μ, 0.4) in (0,1)
Mean (μ)
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

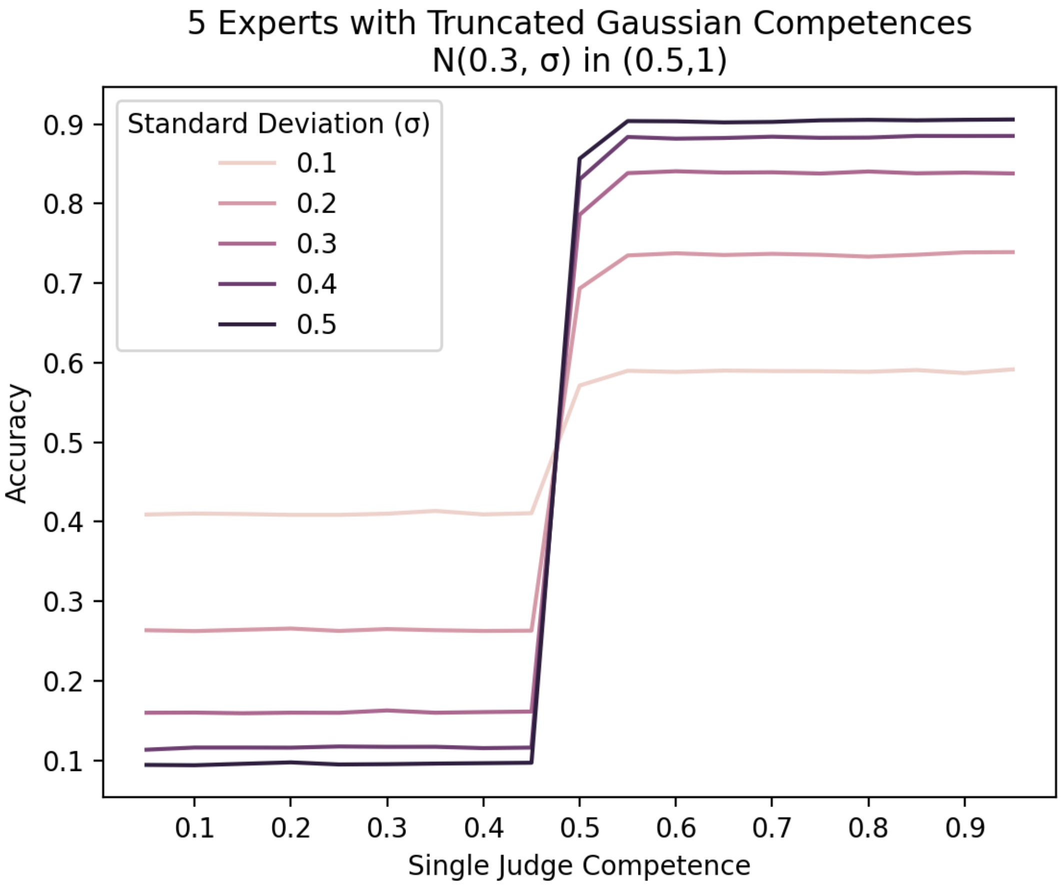

5 Experts with Truncated Gaussian Competences
N(μ, 0.5) in (0,1)
Mean (μ)
0.0
0.2
0.4
0.6
0.8
1.0
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

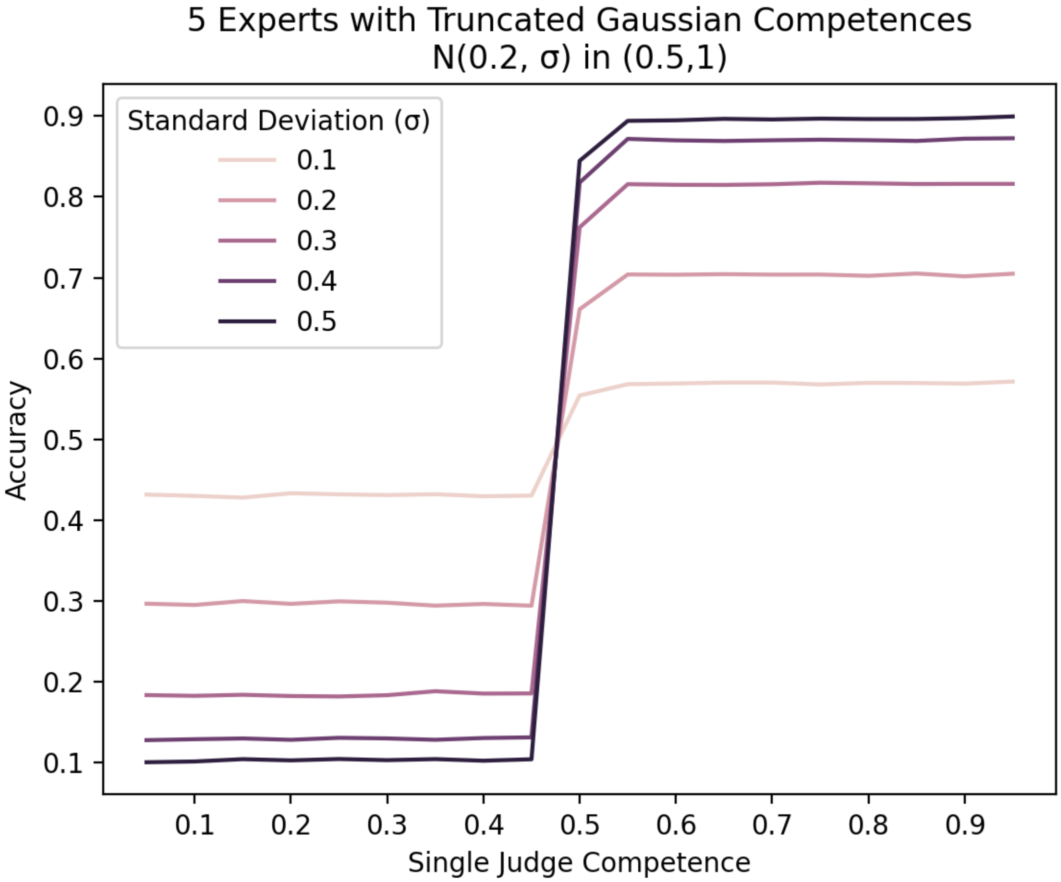
A.4 Normally Distributed competences in $(0.5, 1)$ with 5 Experts

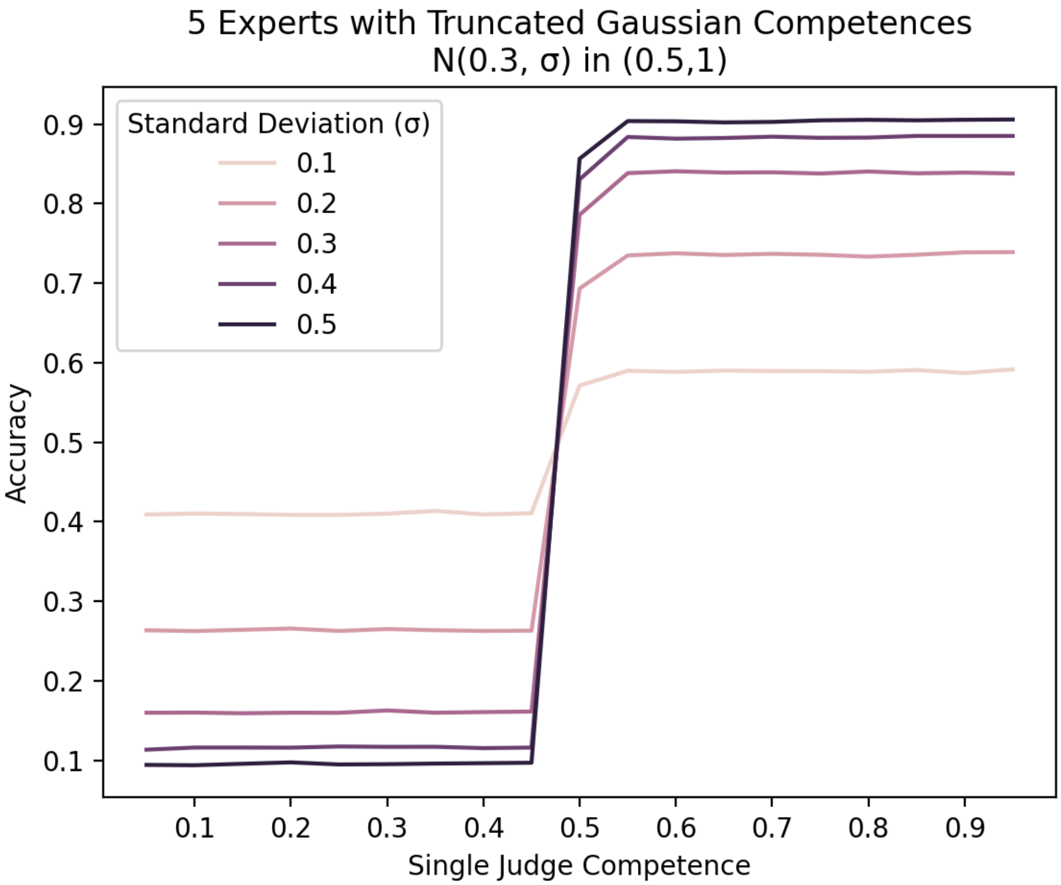


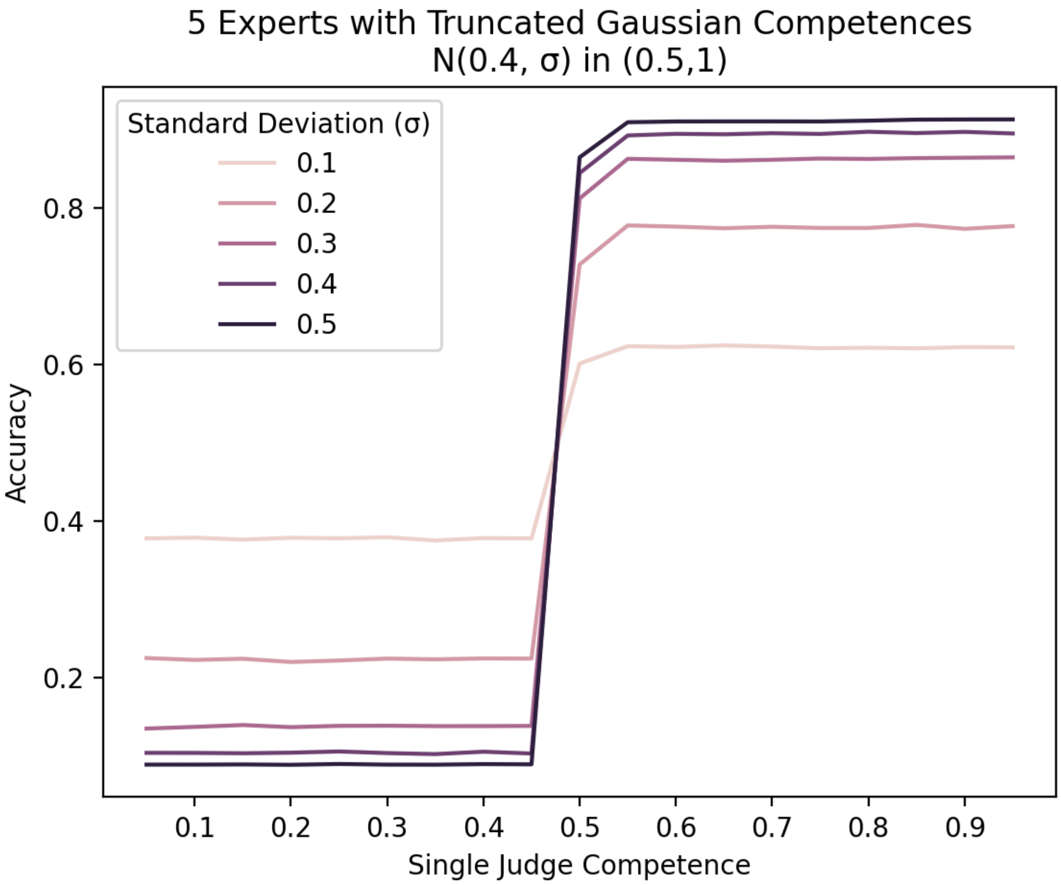

5 Experts with Truncated Gaussian Competences
N(0.1, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

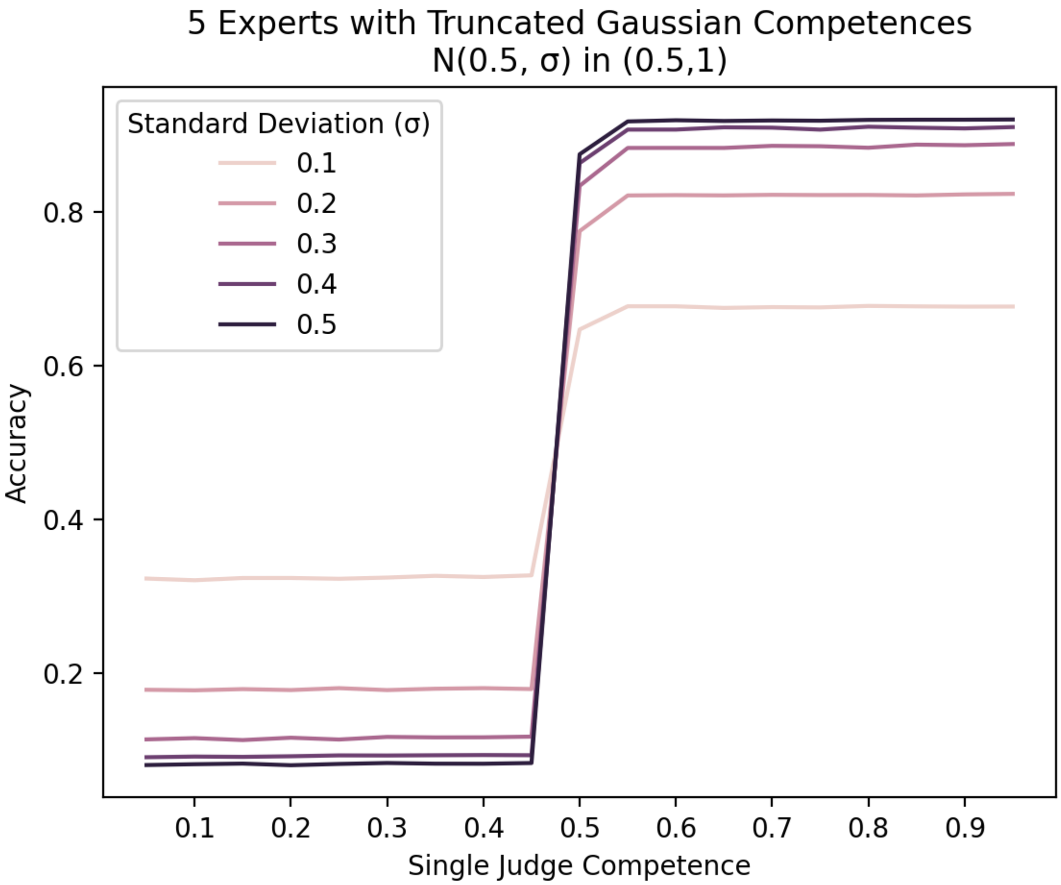

5 Experts with Truncated Gaussian Competences
N(0.2, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

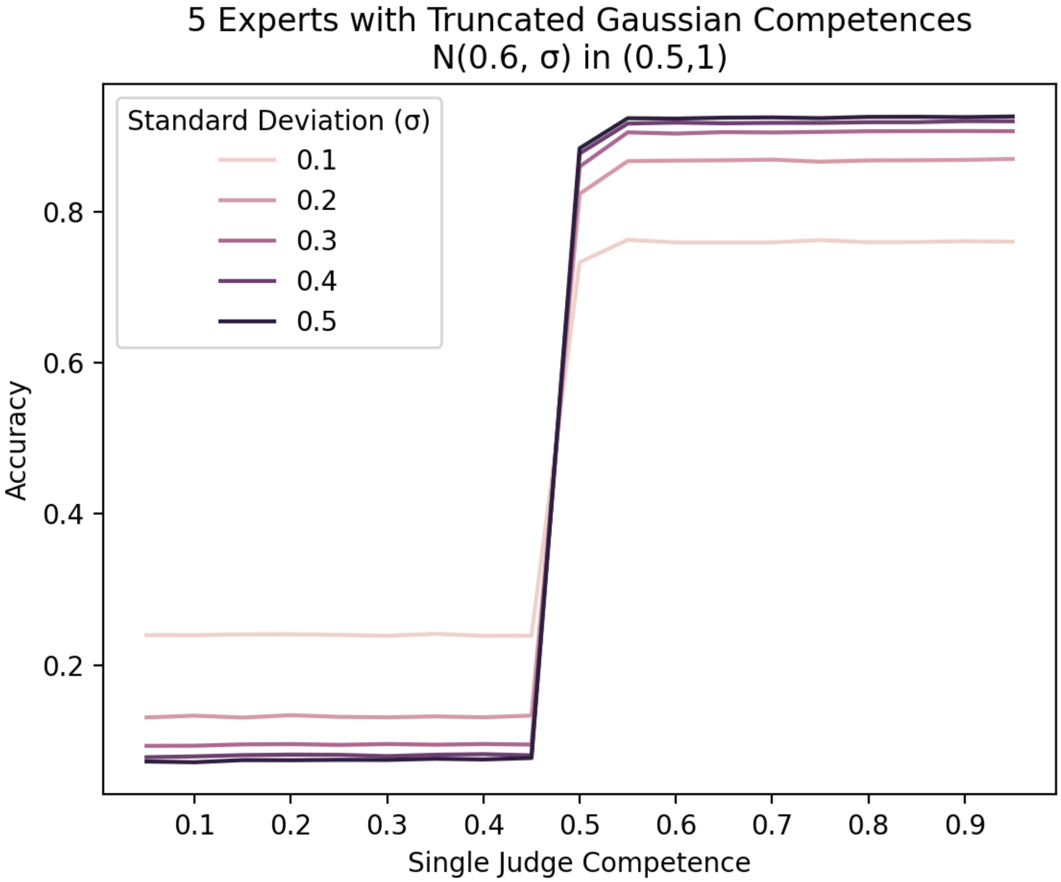

5 Experts with Truncated Gaussian Competences
N(0.3, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

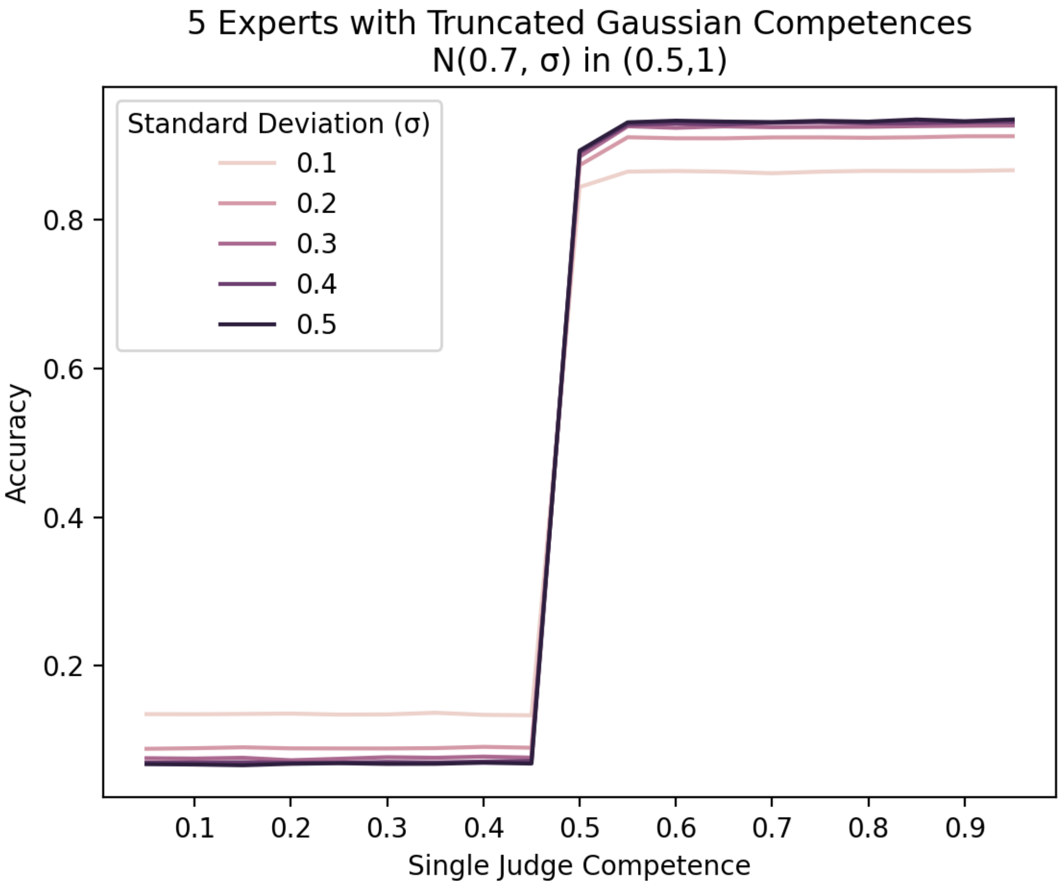

5 Experts with Truncated Gaussian Competences
N(0.2, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

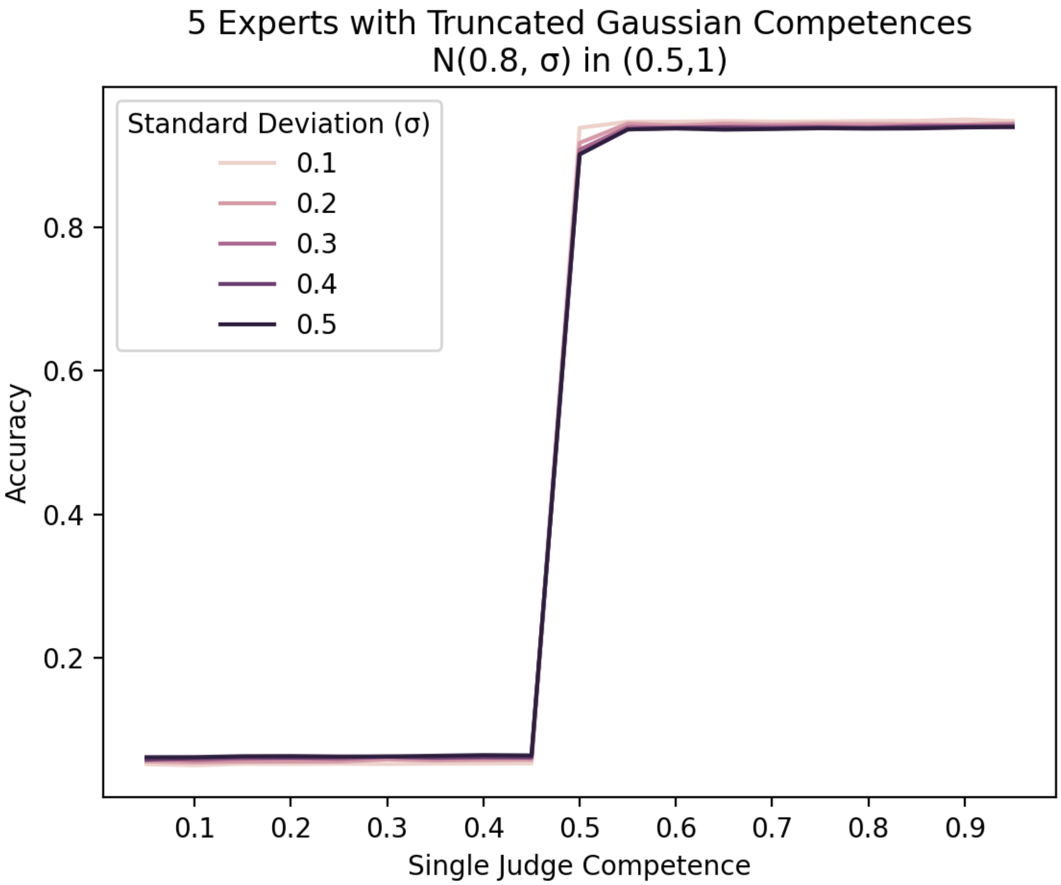

5 Experts with Truncated Gaussian Competences
N(0.3, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

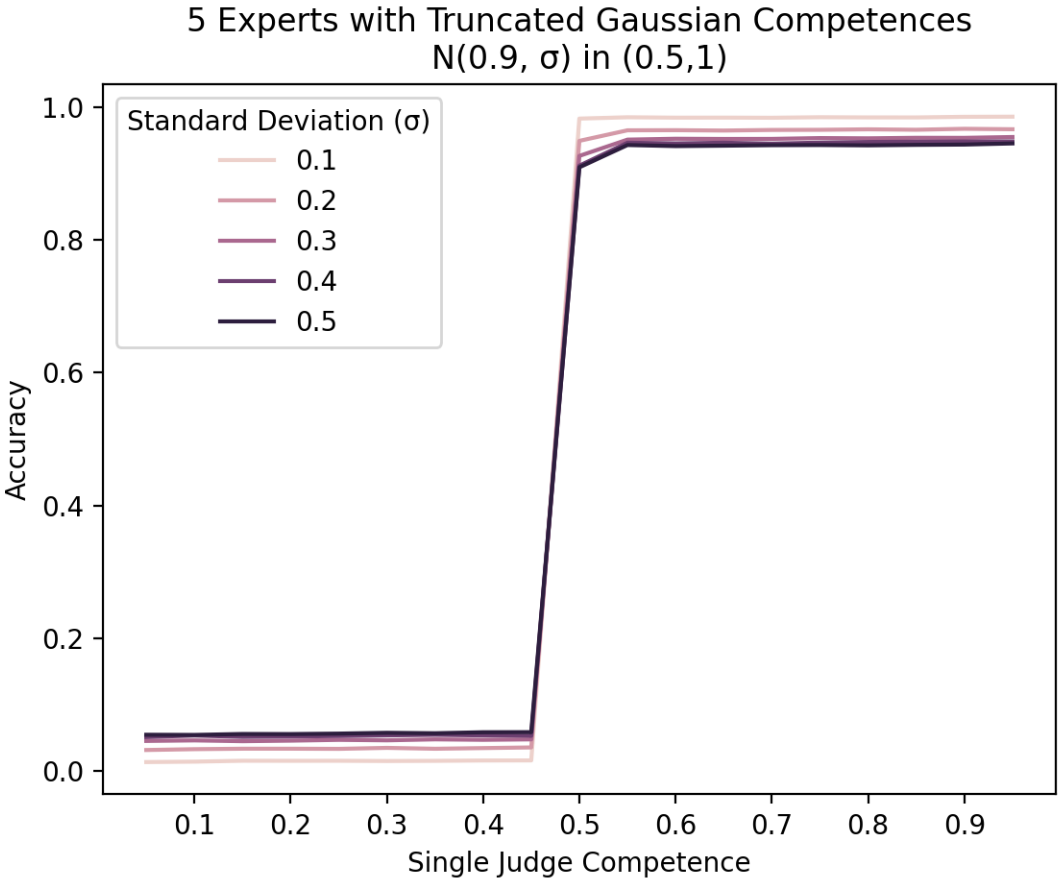

5 Experts with Truncated Gaussian Competences
N(0.4, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

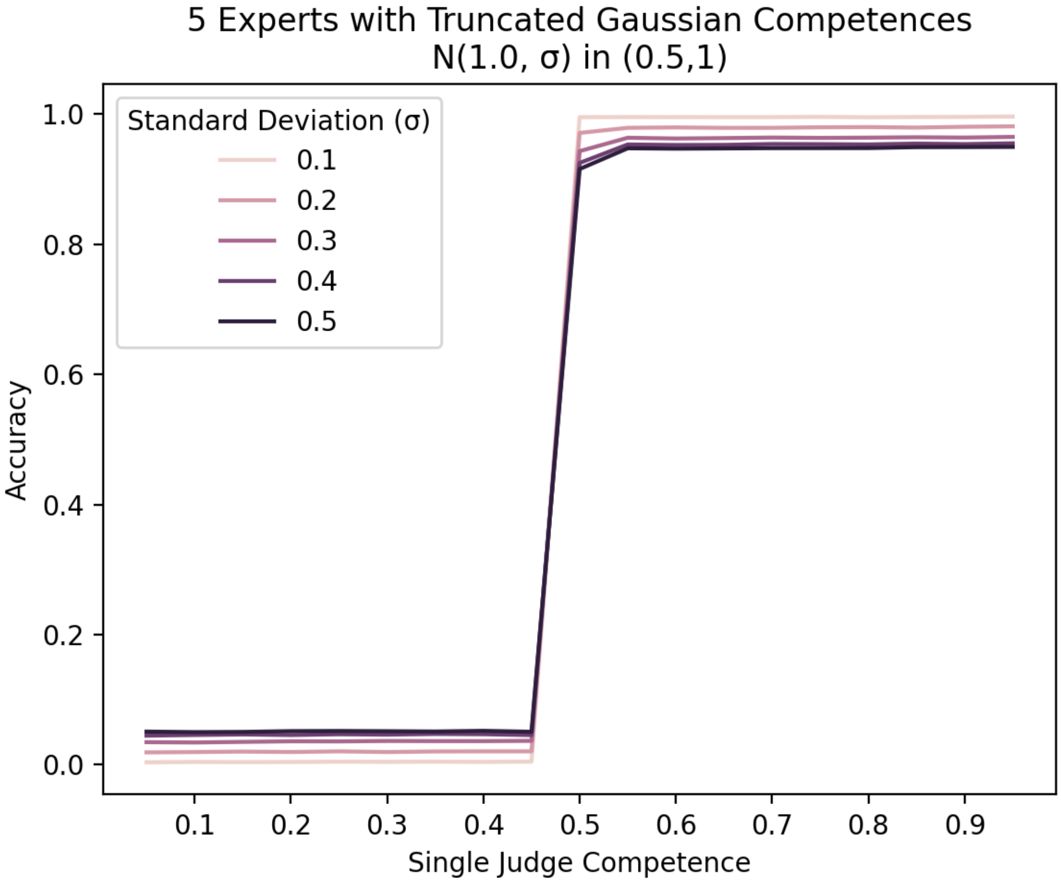

5 Experts with Truncated Gaussian Competences
N(0.5, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.8
0.6
0.4
0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence


5 Experts with Truncated Gaussian Competences
N(0.6, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

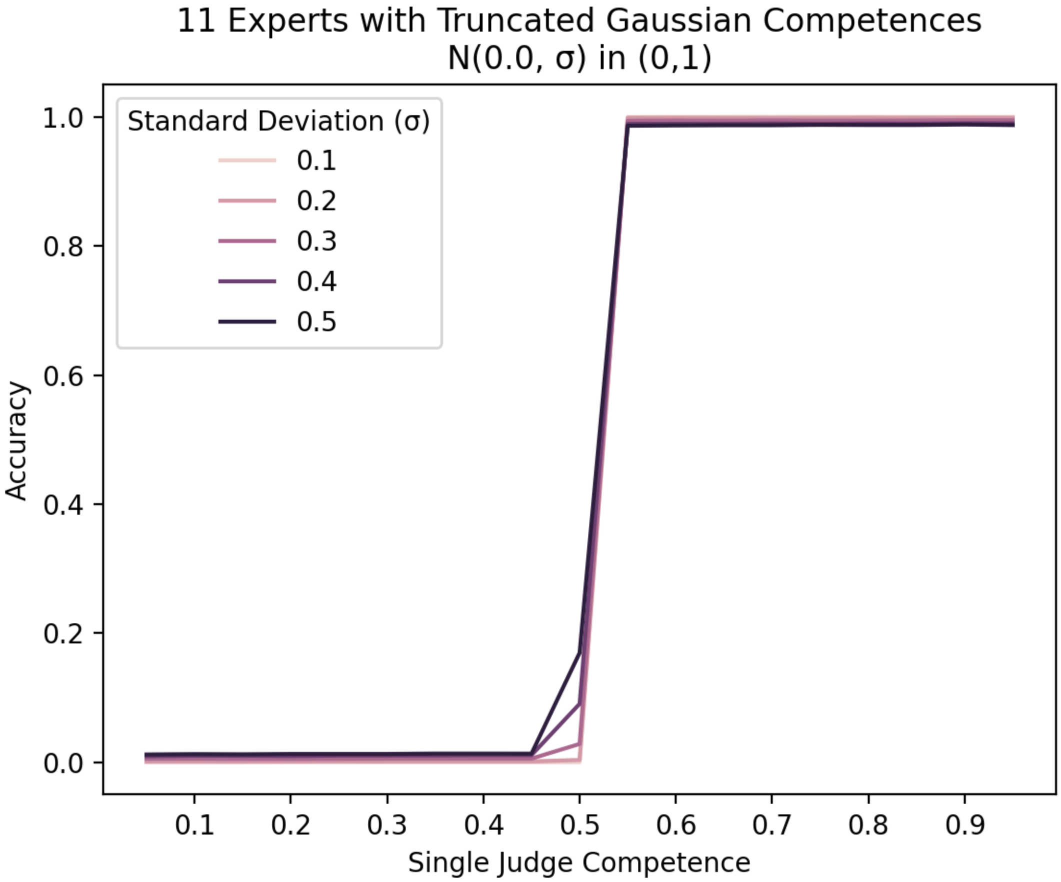

5 Experts with Truncated Gaussian Competences
N(0.7, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

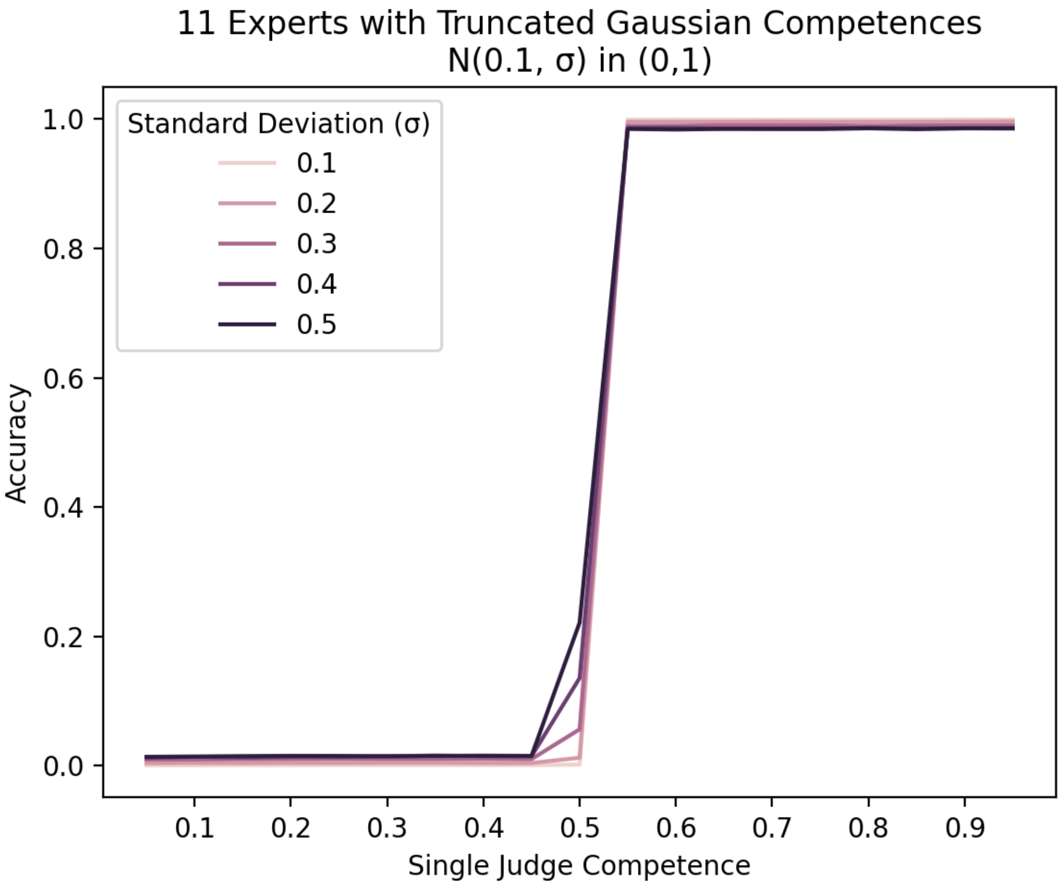

5 Experts with Truncated Gaussian Competences
N(0.8, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.2
0.4
0.6
0.8
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

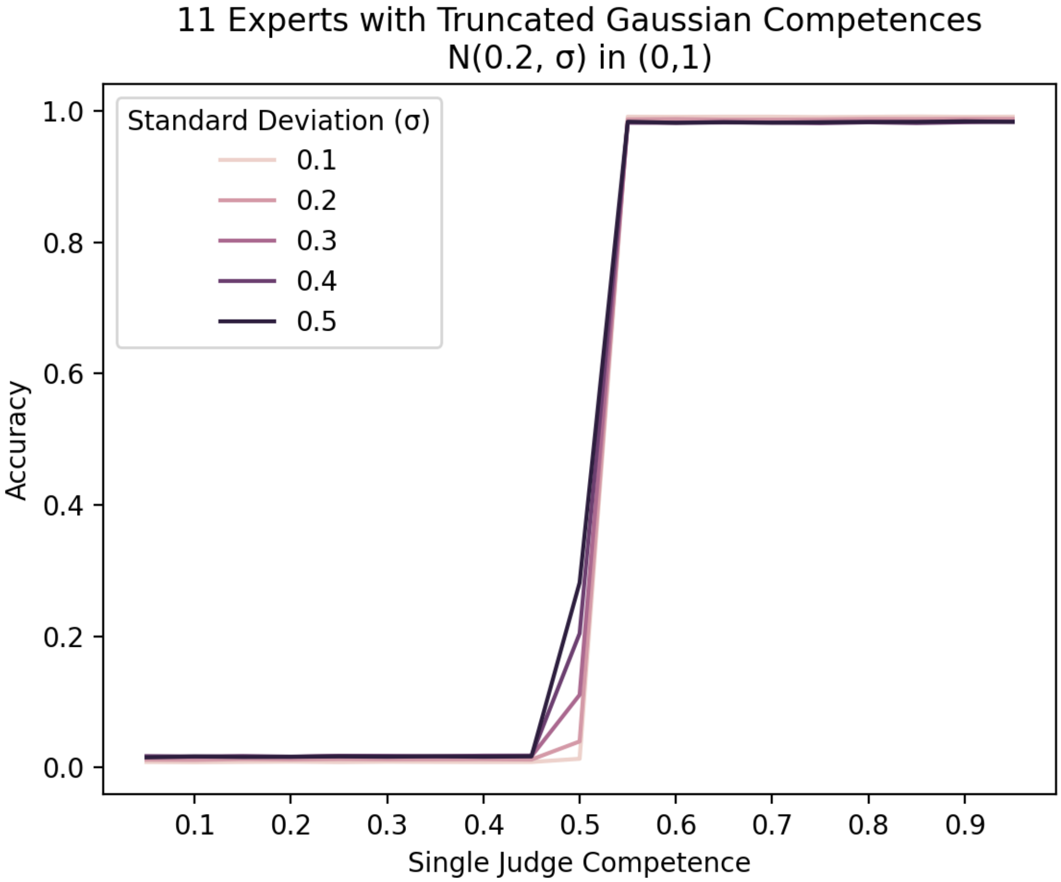

5 Experts with Truncated Gaussian Competences
N(0.9, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

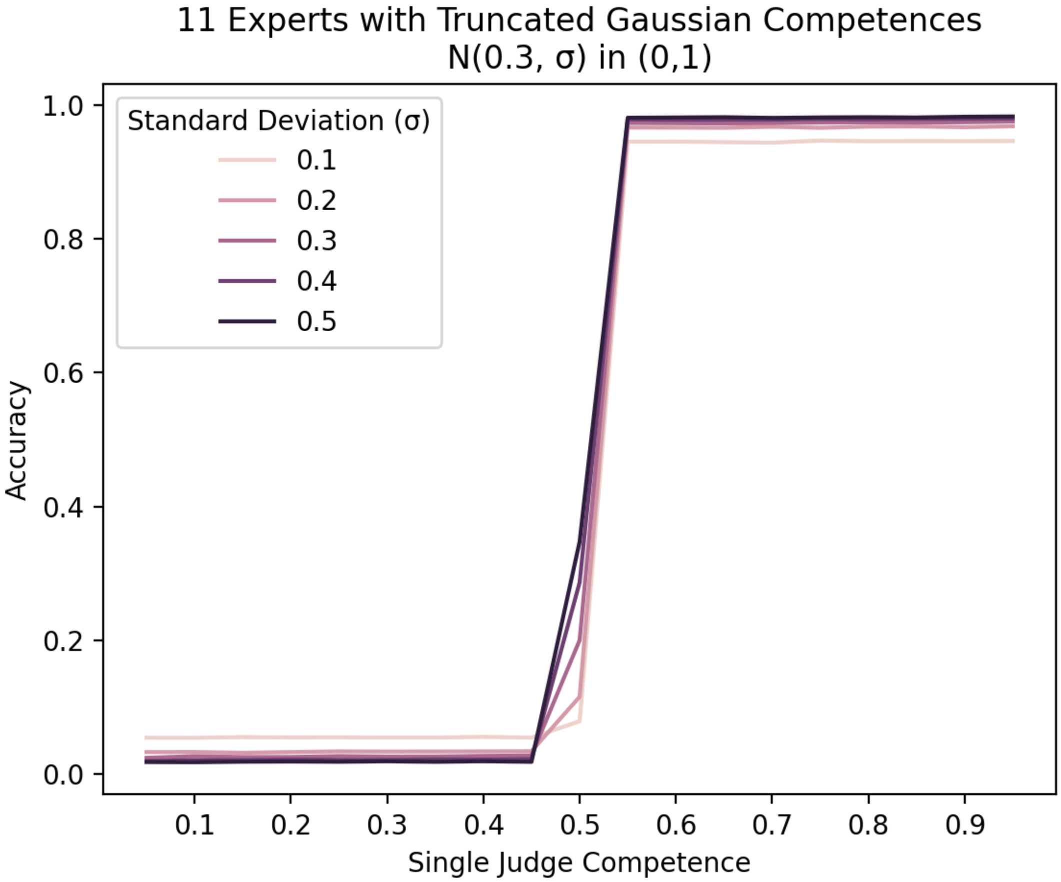

5 Experts with Truncated Gaussian Competences
N(1.0, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

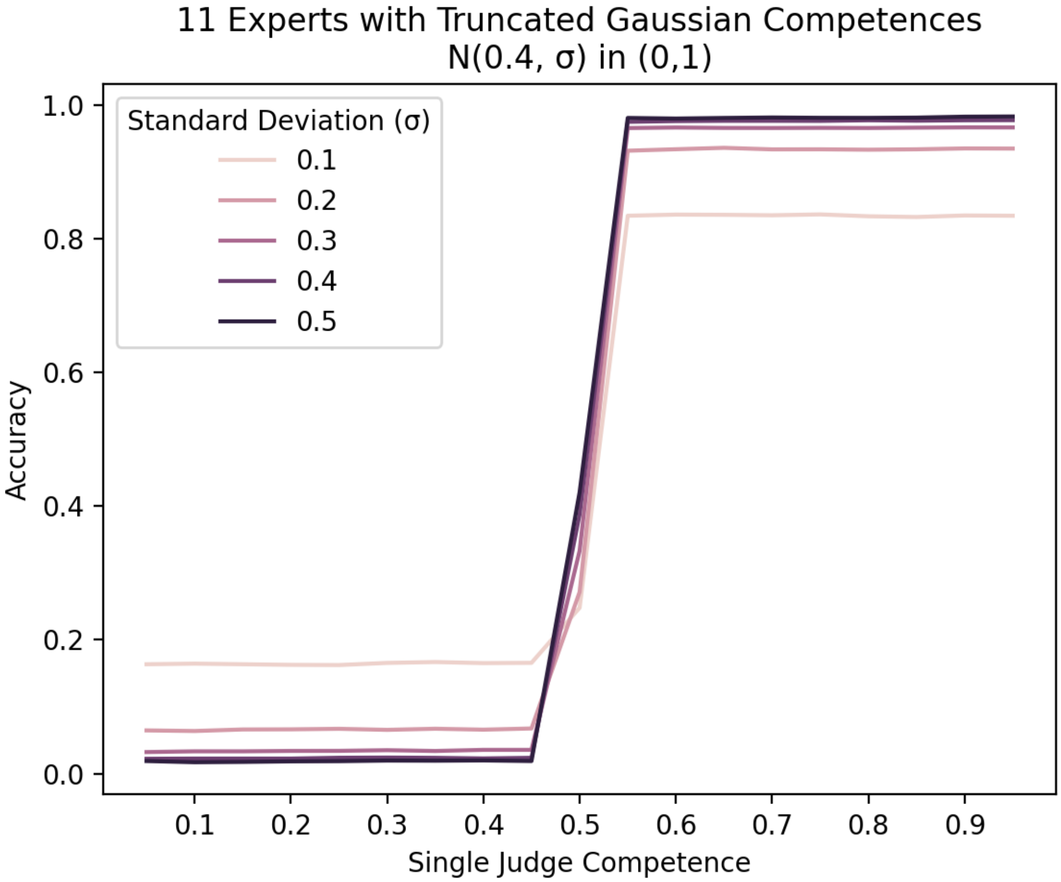
A.5 Normally Distributed competences in $(0, 1)$ with 11 Experts

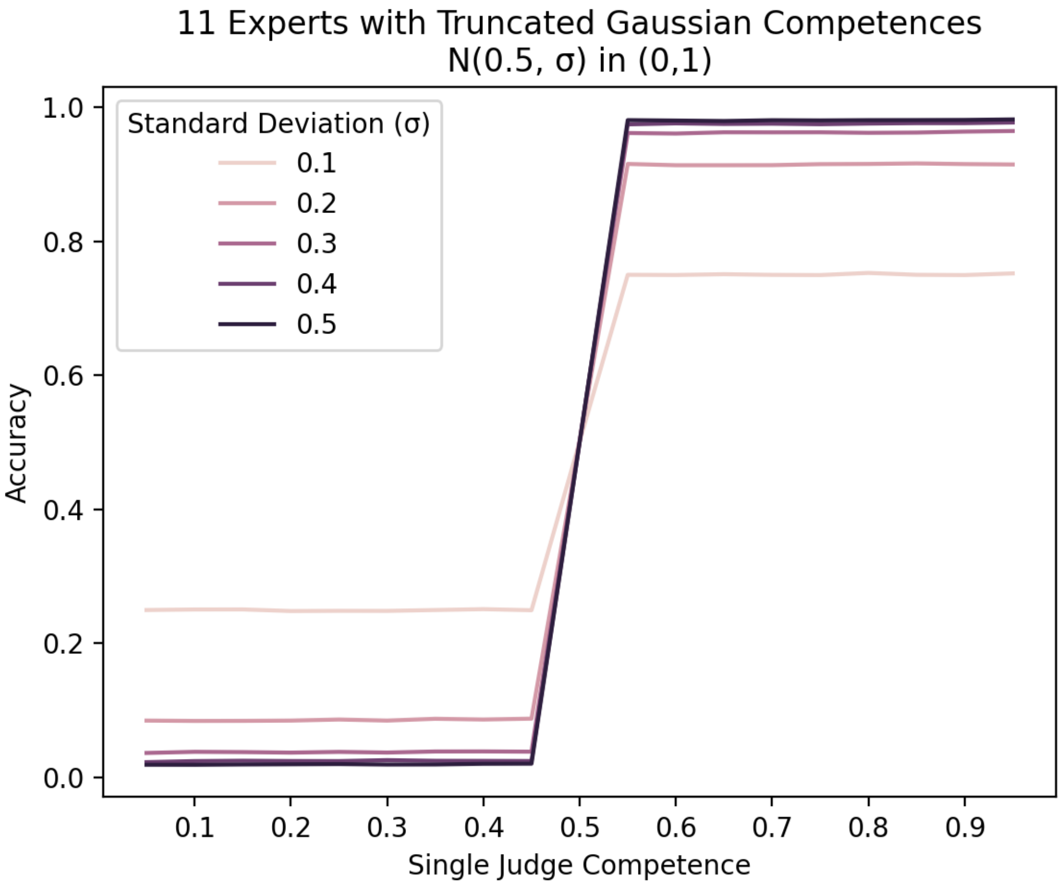


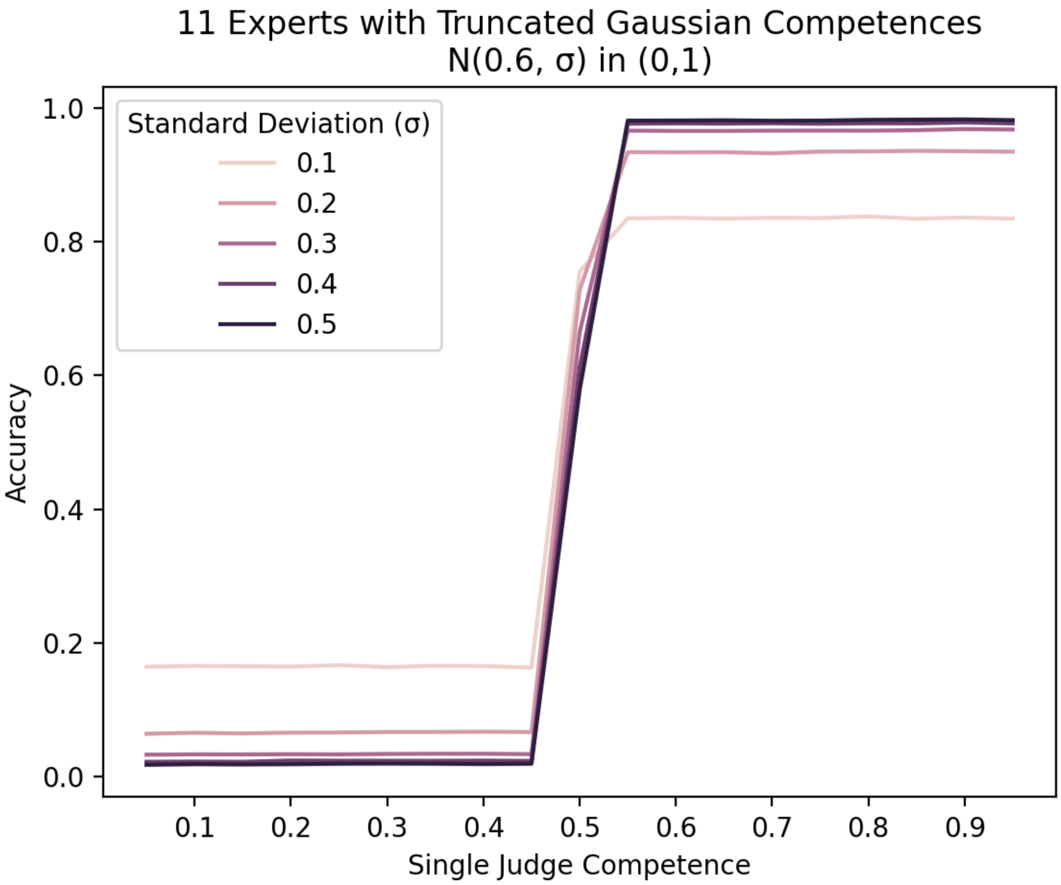

11 Experts with Truncated Gaussian Competences
N(0.1, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

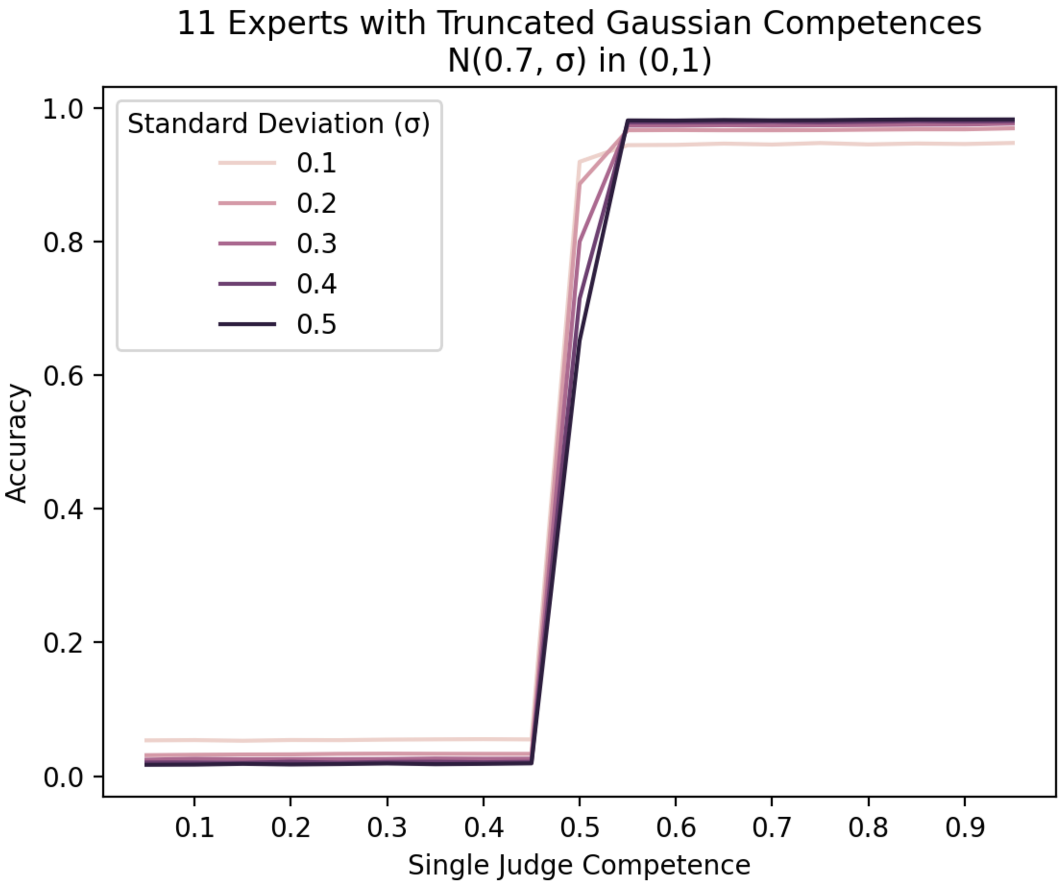

11 Experts with Truncated Gaussian Competences
N(0.2, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

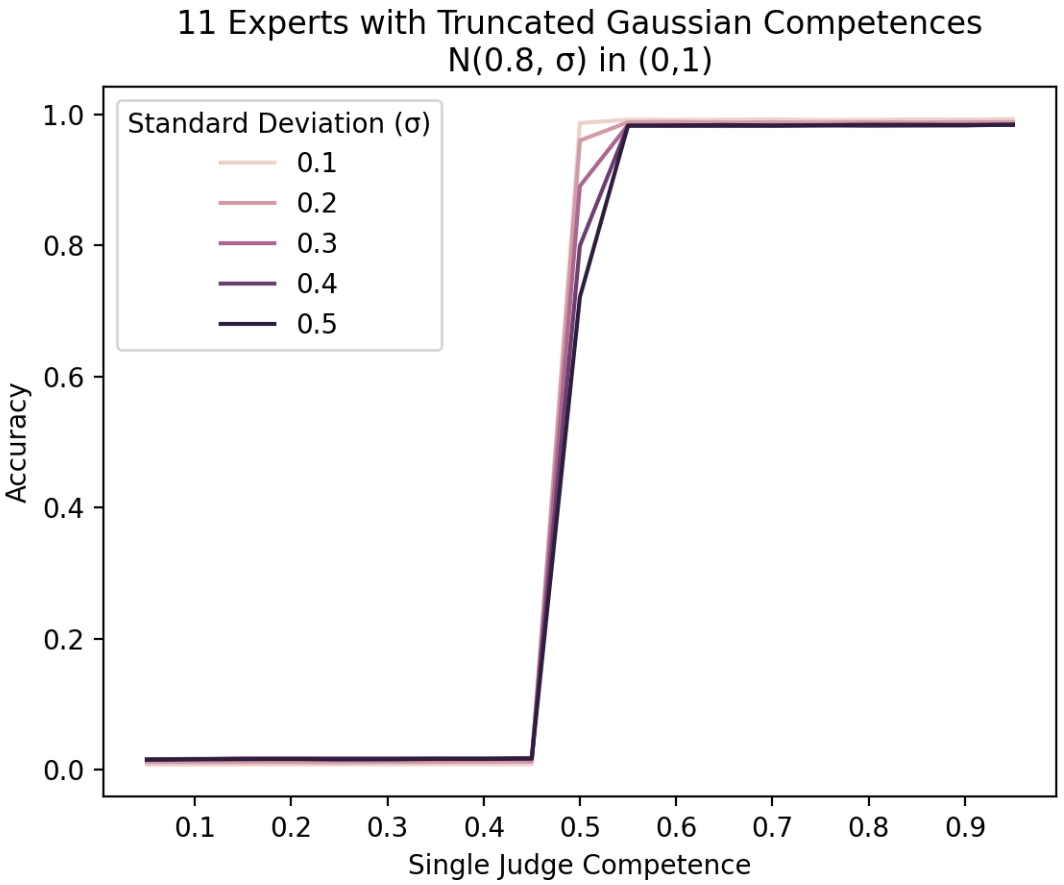

11 Experts with Truncated Gaussian Competences
N(0.3, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

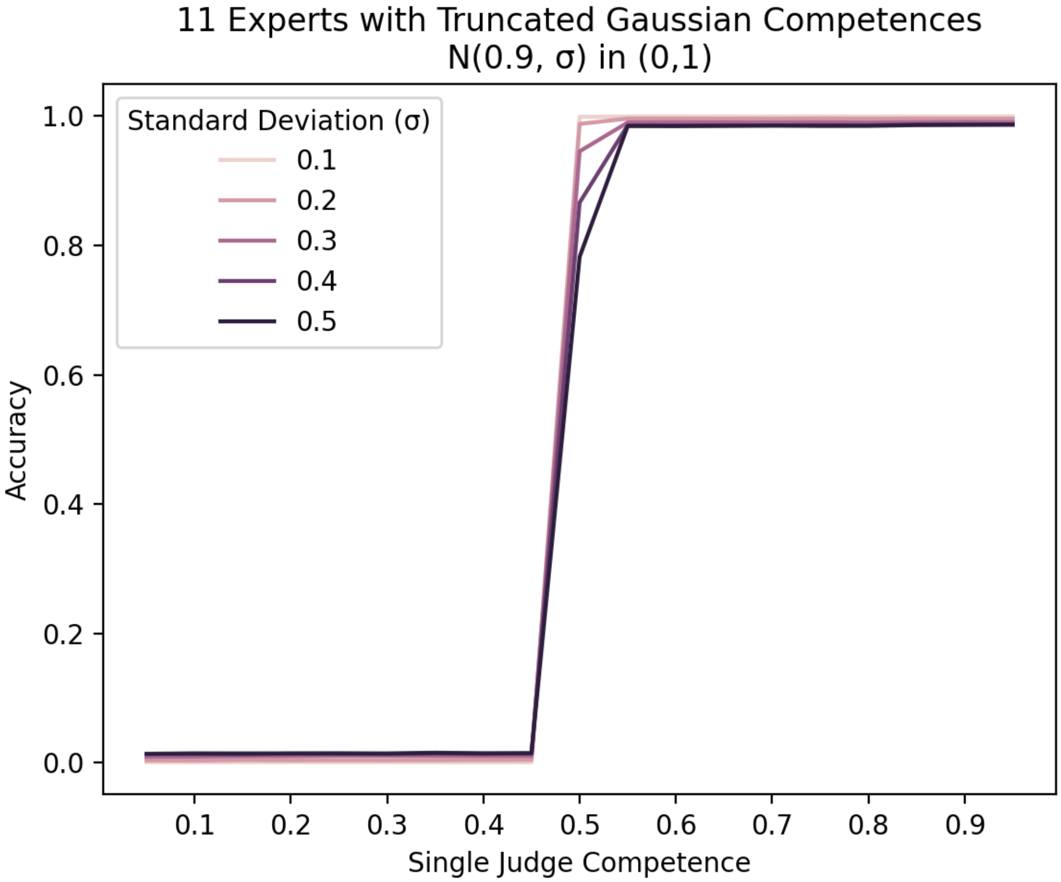

11 Experts with Truncated Gaussian Competences
N(0.4, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

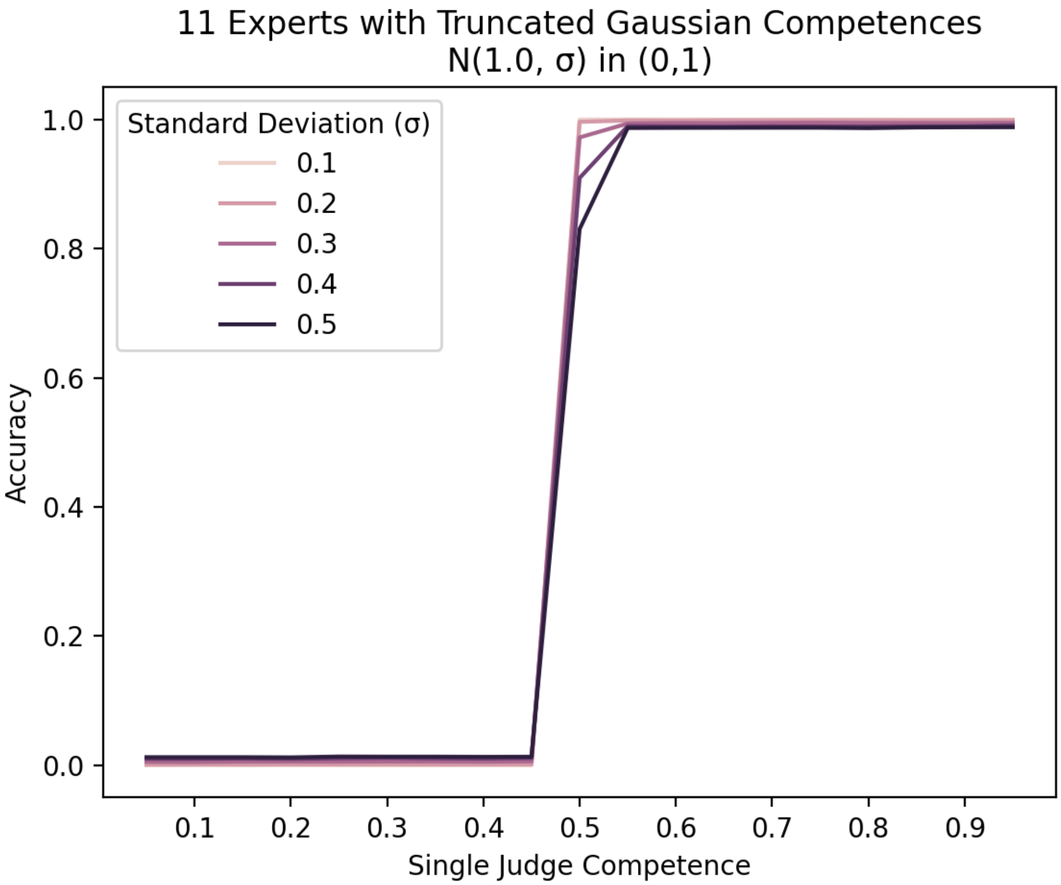

11 Experts with Truncated Gaussian Competences
N(0.5, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


11 Experts with Truncated Gaussian Competences
N(0.6, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

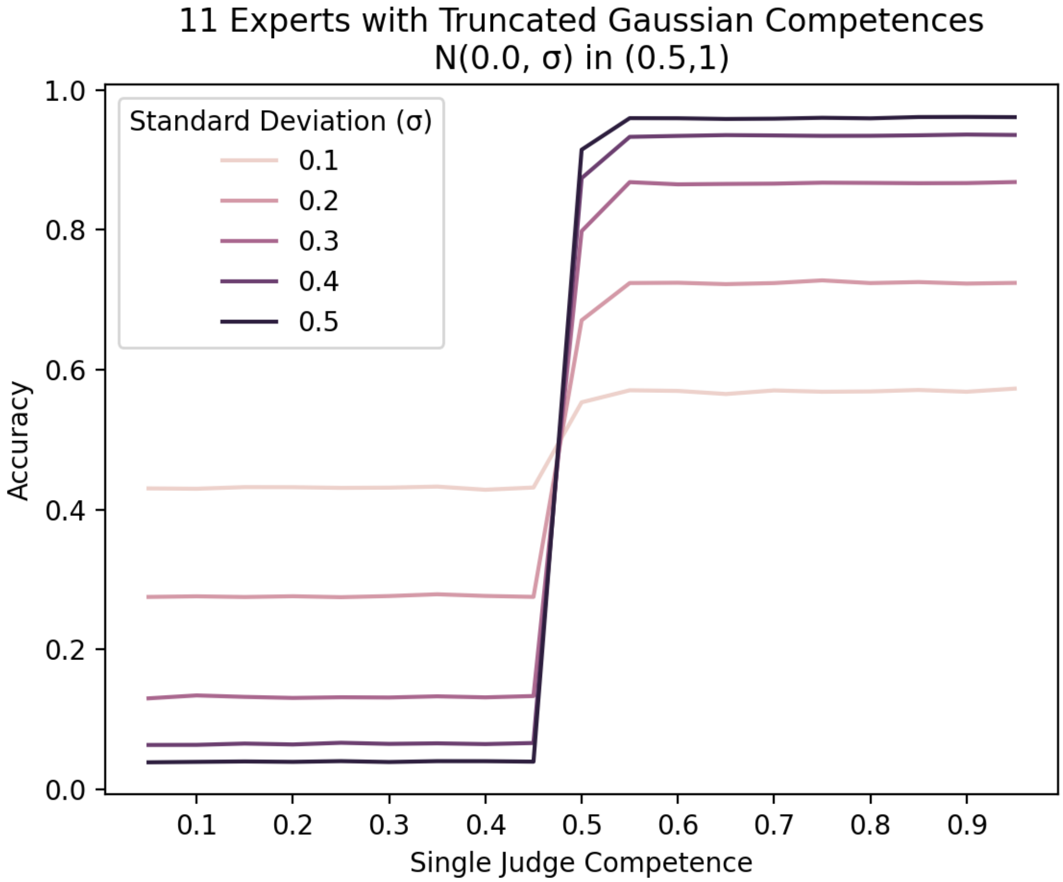

11 Experts with Truncated Gaussian Competences
N(0.7, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

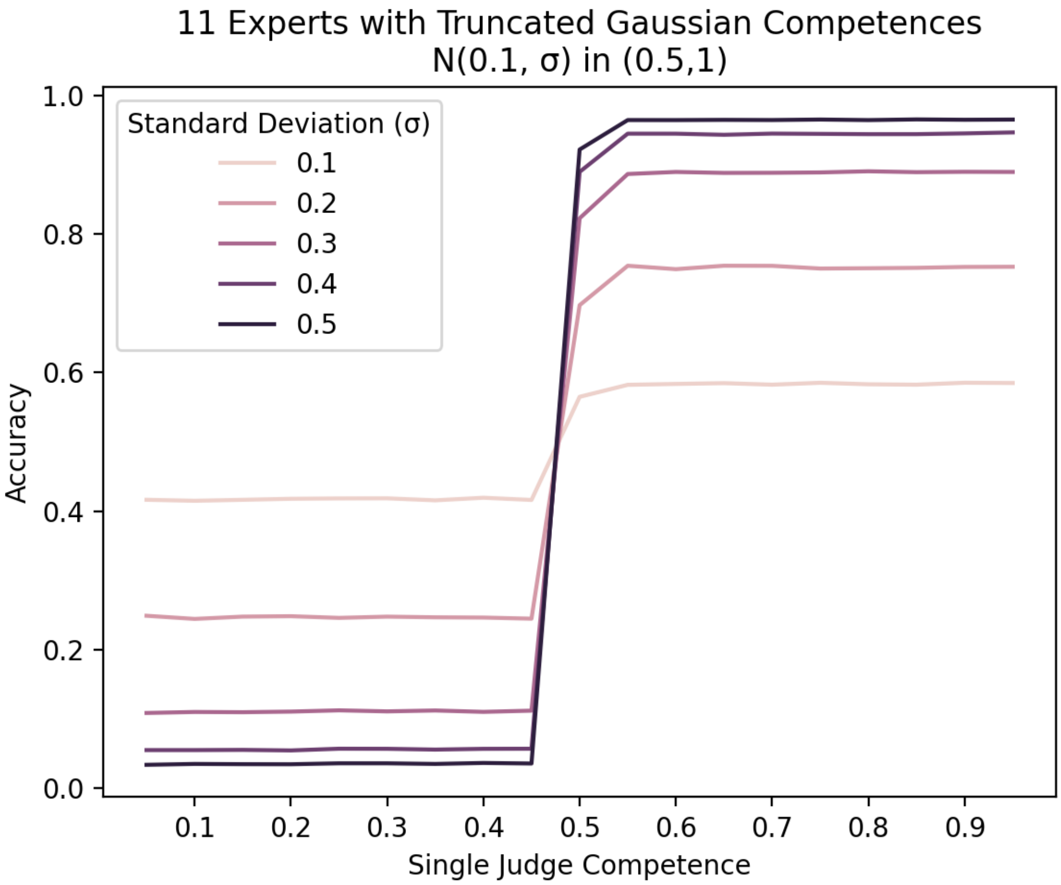

11 Experts with Truncated Gaussian Competences
N(0.8, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

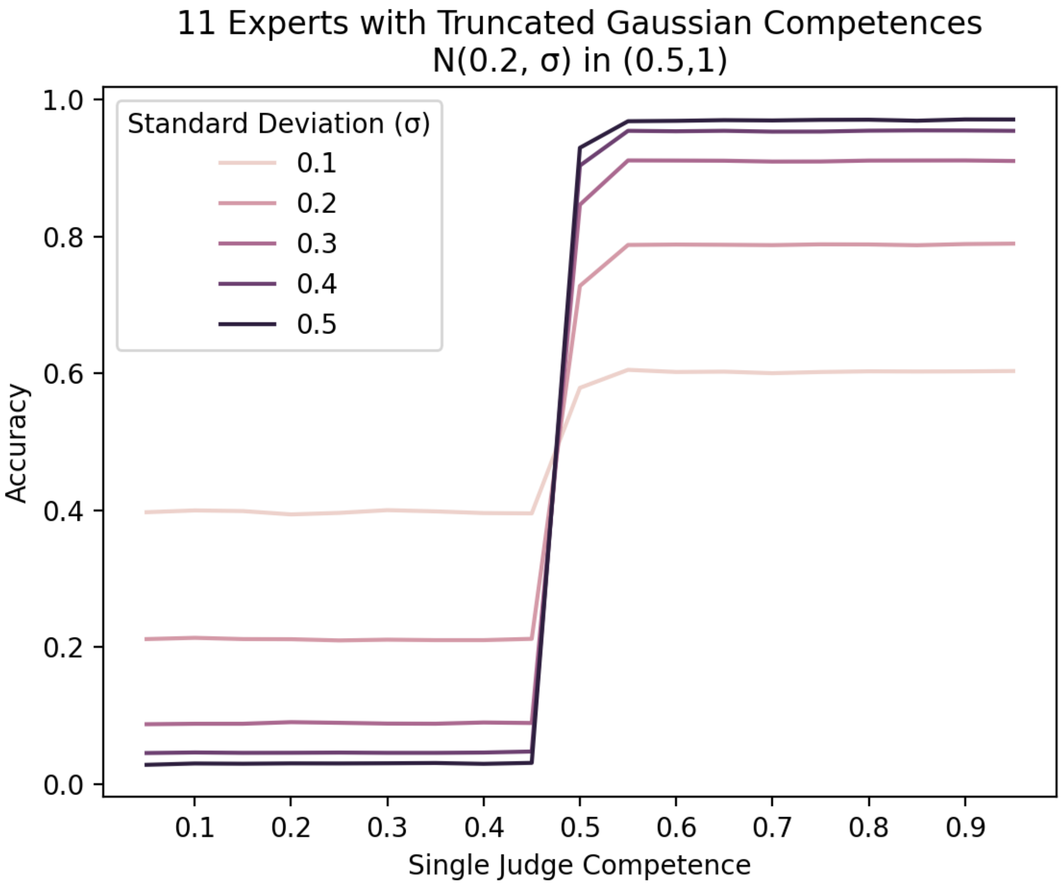

11 Experts with Truncated Gaussian Competences
N(0.9, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

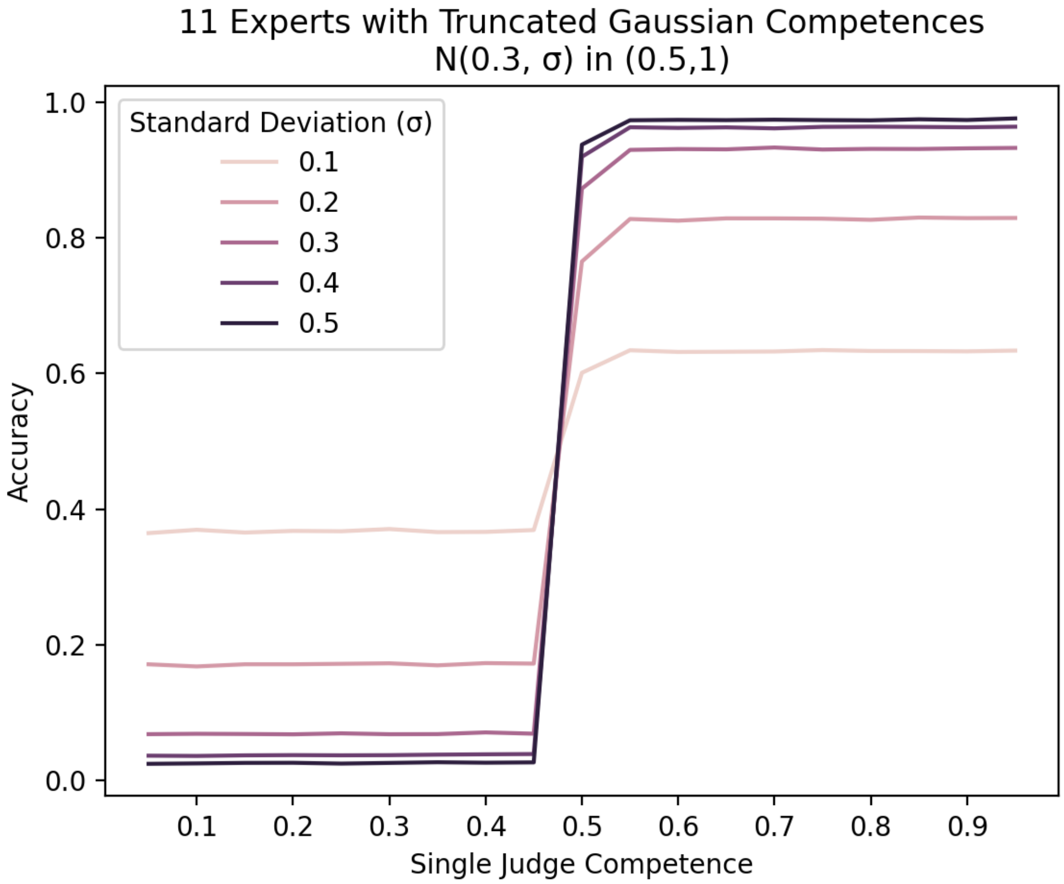

11 Experts with Truncated Gaussian Competences
N(1.0, σ) in (0,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

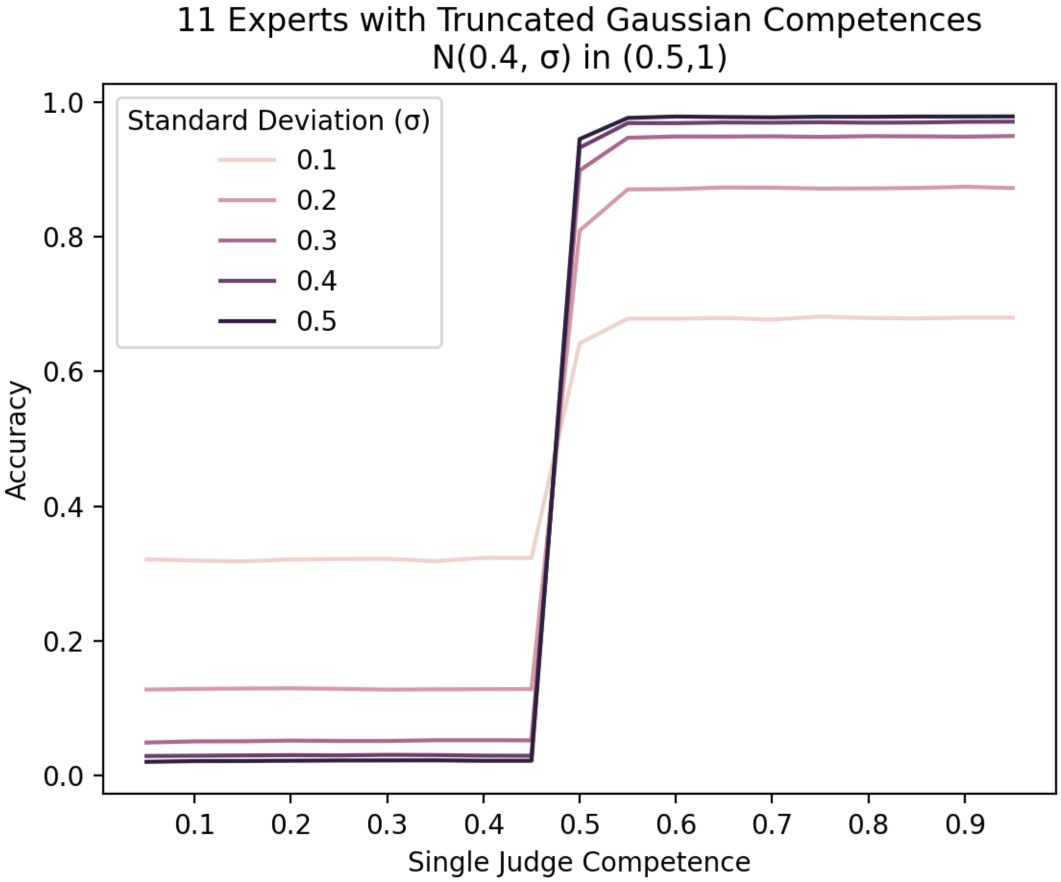
A.6 Normally Distributed competences in $(0.5, 1)$ with 11 Experts

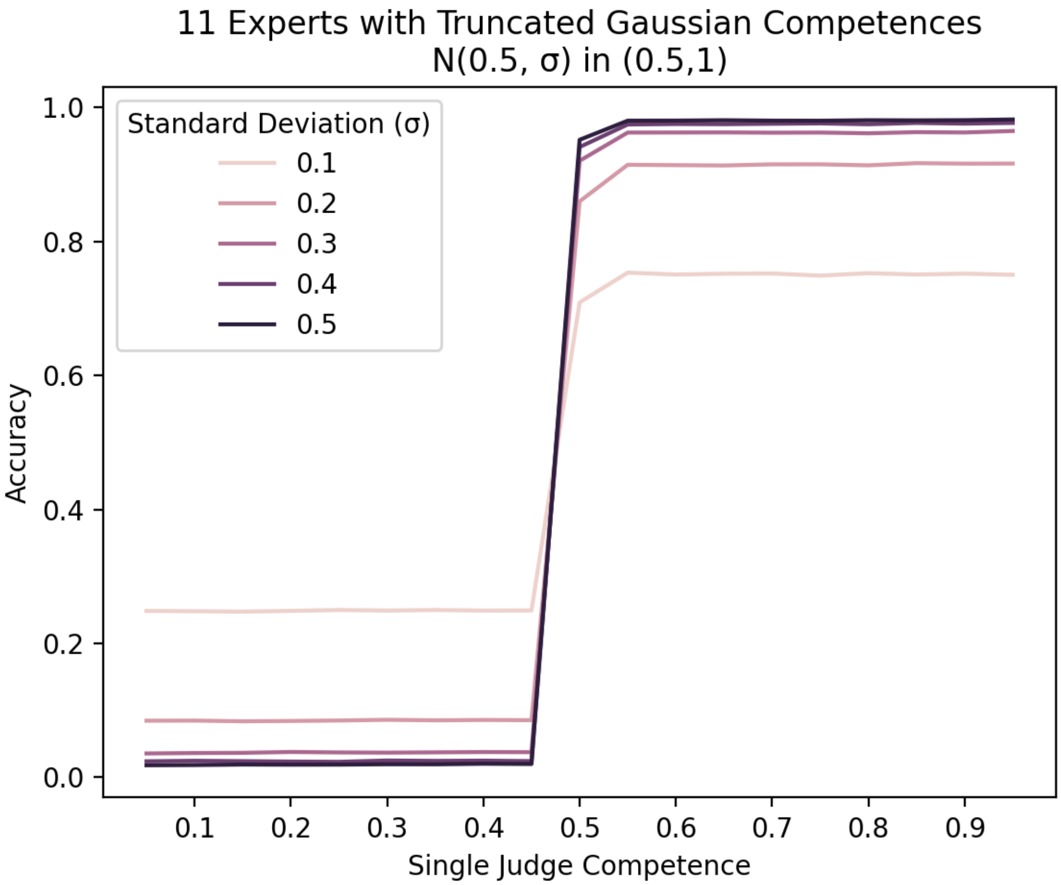


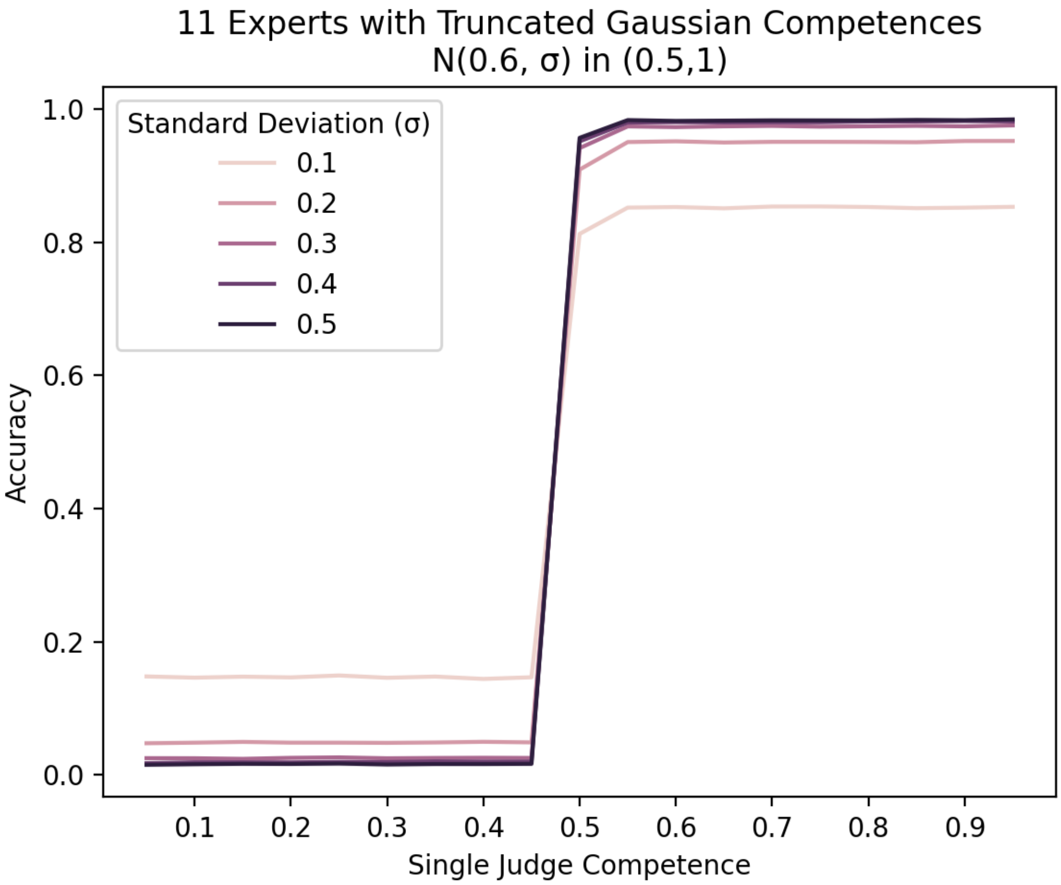

11 Experts with Truncated Gaussian Competences
N(0.1, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

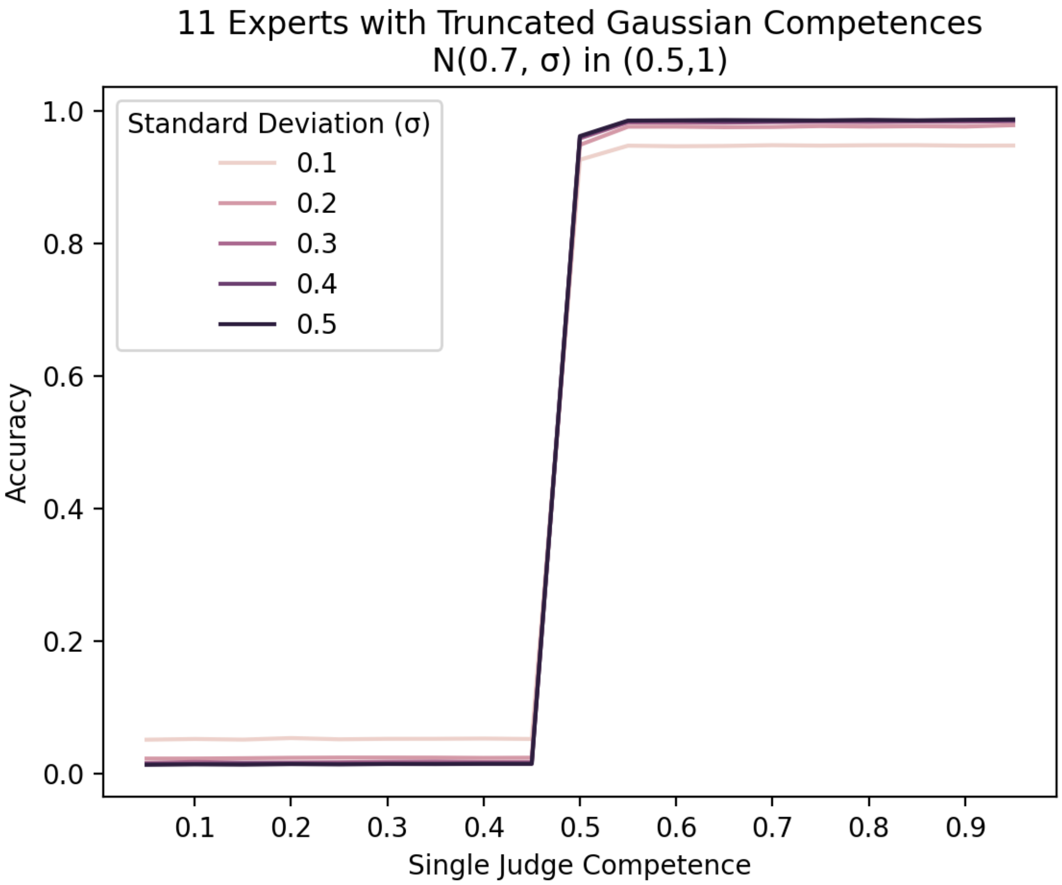

11 Experts with Truncated Gaussian Competences
N(0.2, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

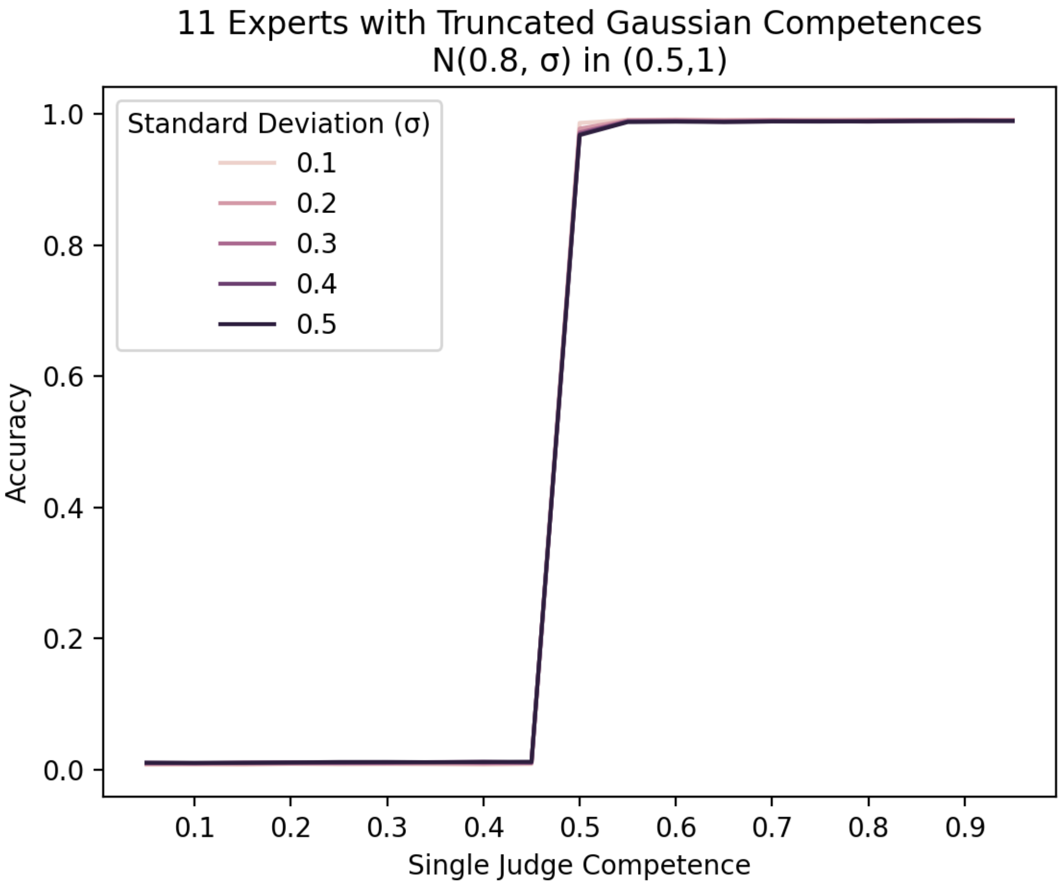

11 Experts with Truncated Gaussian Competences
N(0.3, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

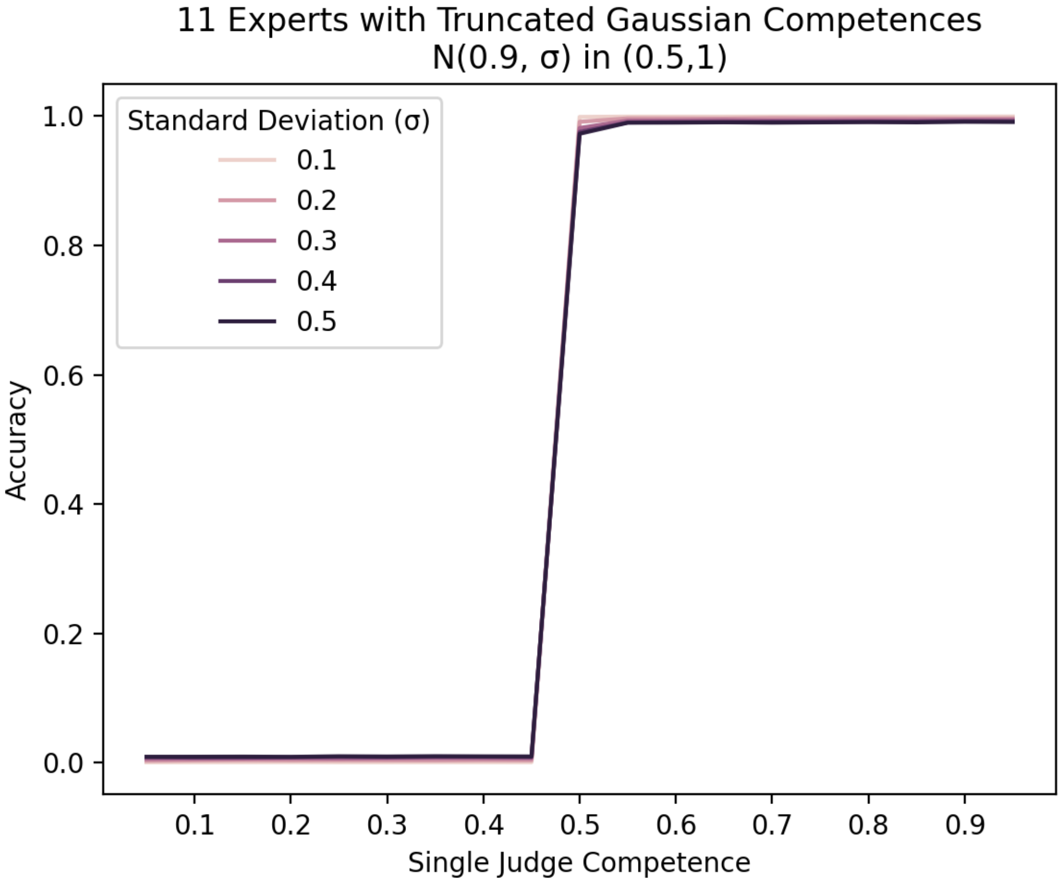

11 Experts with Truncated Gaussian Competences
N(0.4, σ) in (0.5,1)
Standard Deviation (σ)
— 0.1
— 0.2
— 0.3
— 0.4
— 0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

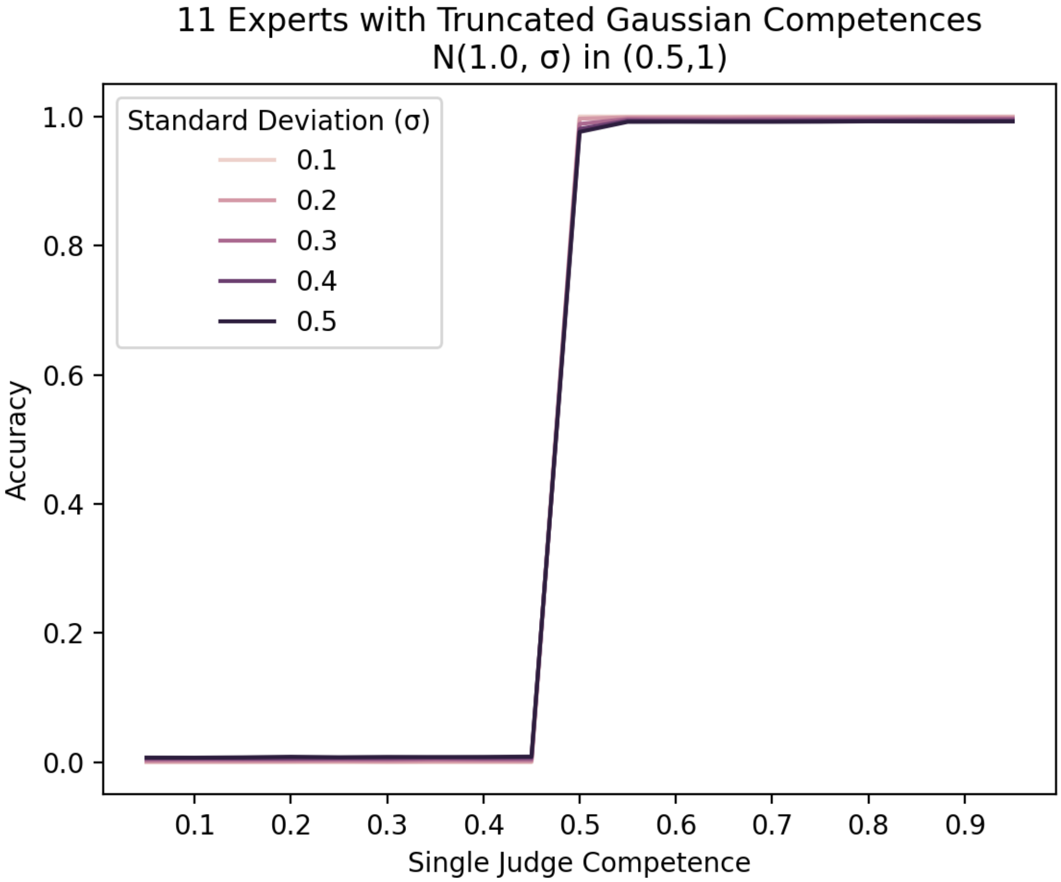

11 Experts with Truncated Gaussian Competences
N(0.5, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


11 Experts with Truncated Gaussian Competences
N(0.6, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


11 Experts with Truncated Gaussian Competences
N(0.7, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

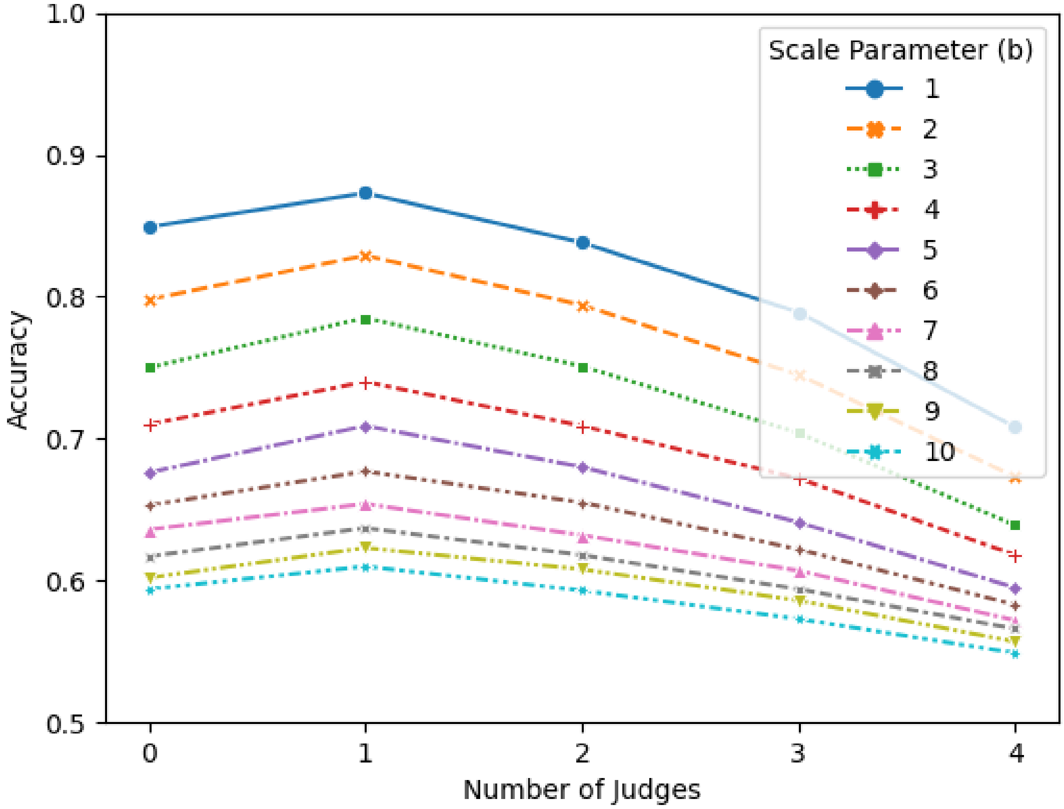

11 Experts with Truncated Gaussian Competences
N(0.8, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Single Judge Competence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

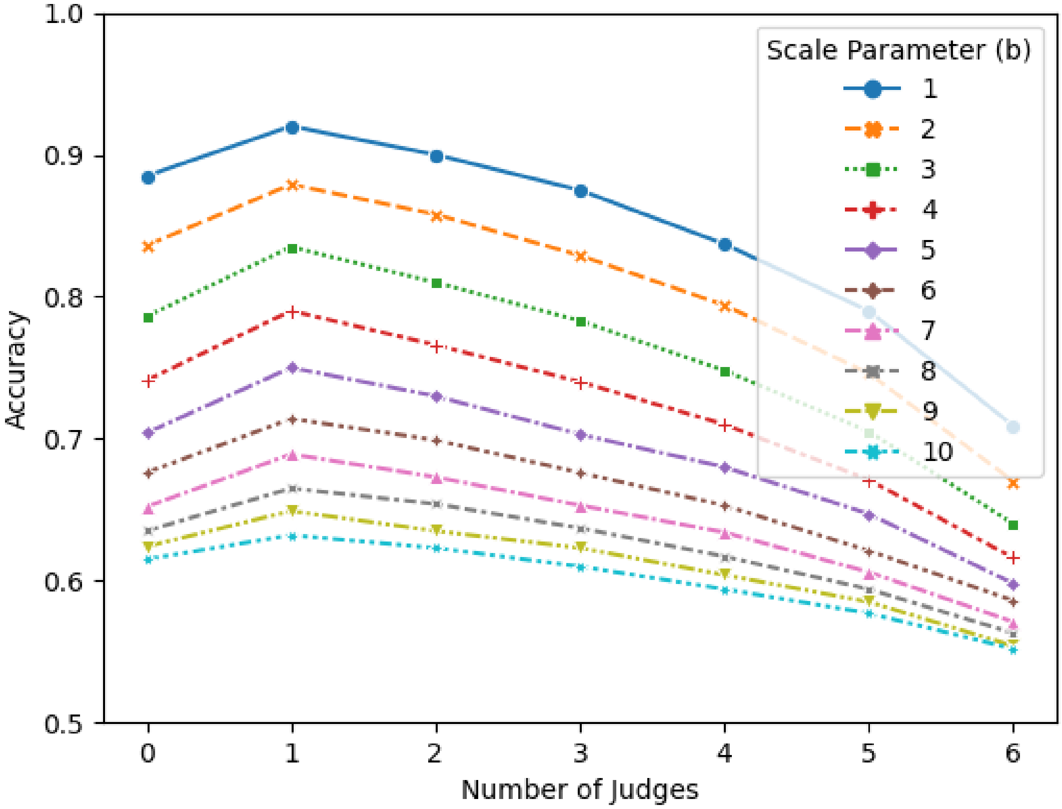

11 Experts with Truncated Gaussian Competences
N(0.9, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

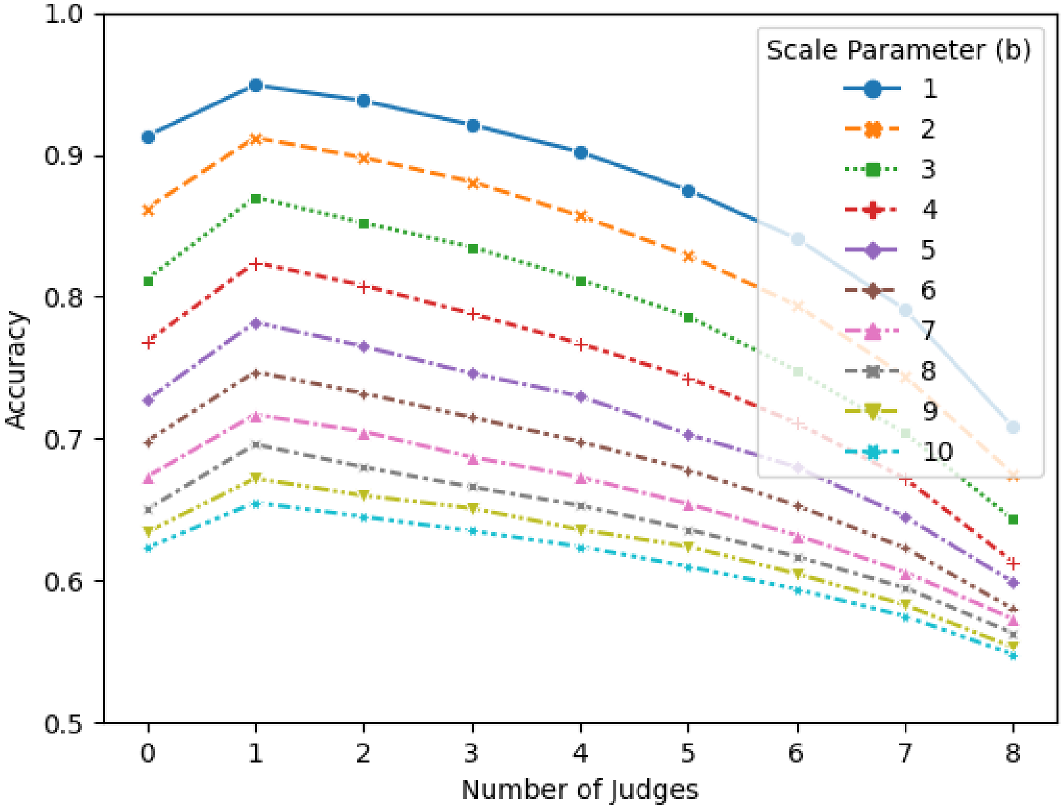

11 Experts with Truncated Gaussian Competences
N(1.0, σ) in (0.5,1)
Standard Deviation (σ)
0.1
0.2
0.3
0.4
0.5
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Single Judge Competence

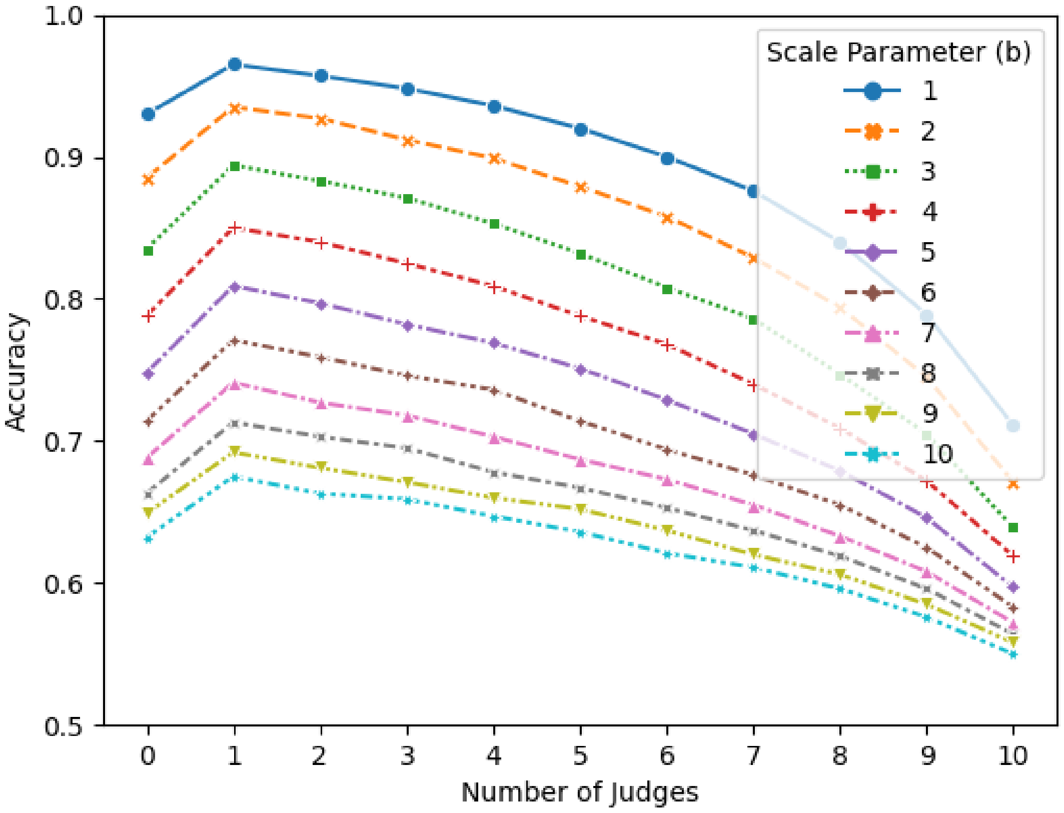
B PARTITIONING AGENTS INTO ROLES

B.1 Exponentially Distributed competences in $(0.5, 1)$

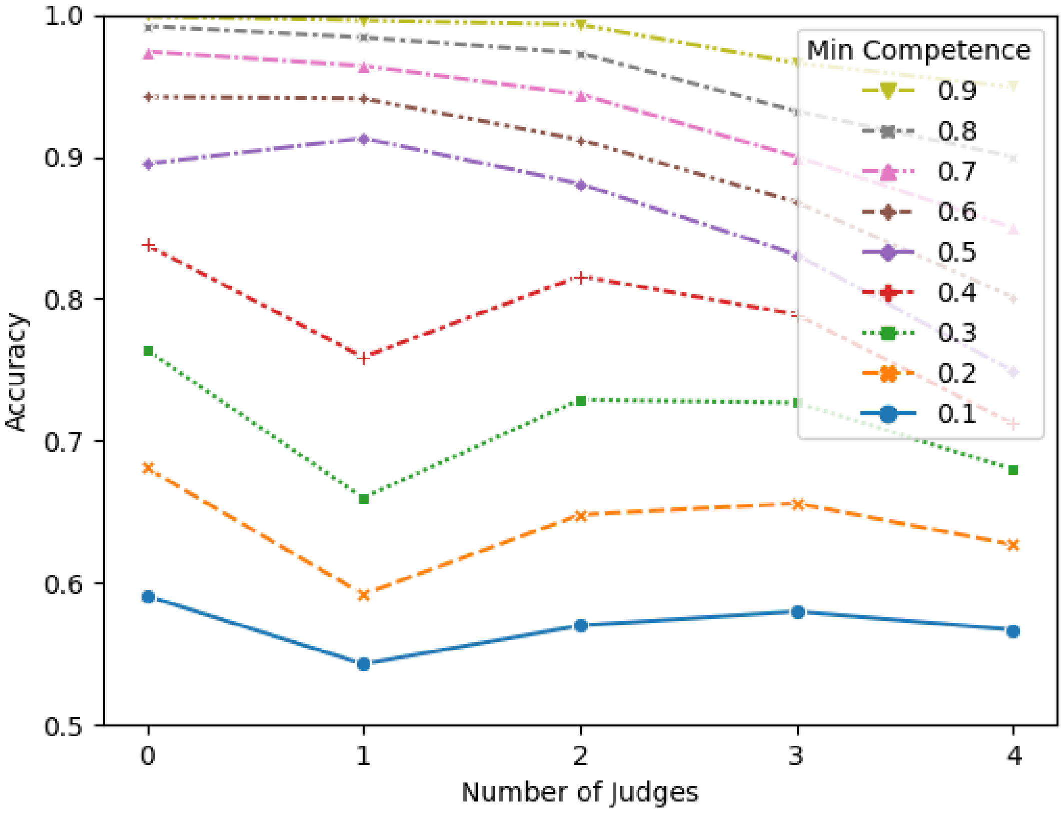


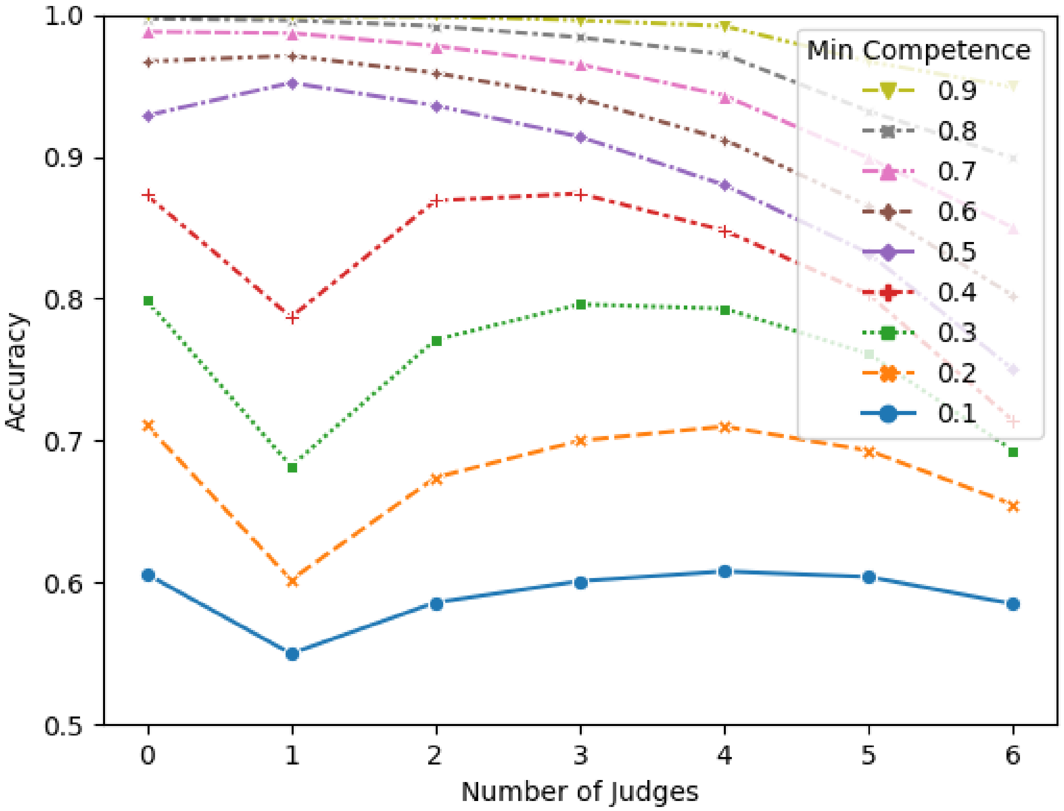

Scale Parameter (b)
1
2
3
4
5
6
7
8
9
10
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
Number of Judges

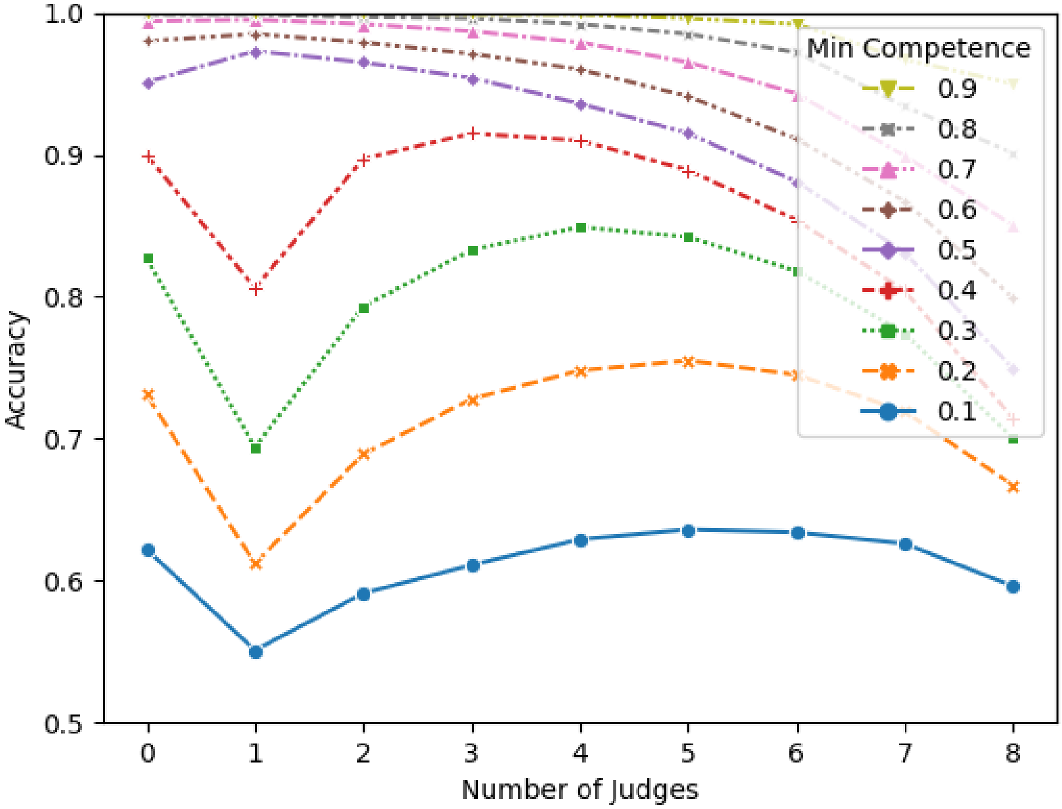

Scale Parameter (b)
1
2
3
4
5
6
7
8
9
10
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
Number of Judges
0
1
2
3
4
5
6
7
8

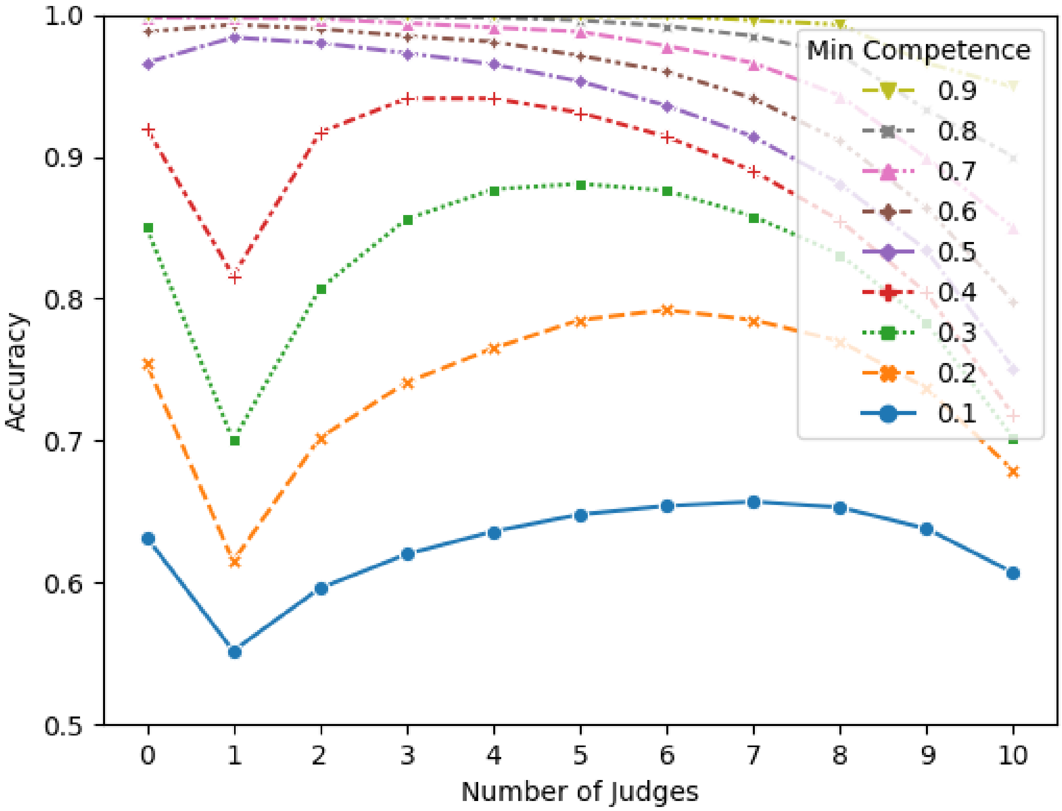

Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
7
8
9
10
Number of Judges
Scale Parameter (b)
1
2
3
4
5
6
7
8
9
10

B.2 Uniformly Distributed competences




Min Competence
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
Number of Judges
0
1
2
3
4
5
6


Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
7
8
Number of Judges
Min Competence
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1


Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
7
8
9
10
Number of Judges
Min Competence
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1