Multidimensional Information and Rational Inattention*

[Click here for the latest version]

Peicong (Keri) Hu†

November 22, 2020

Abstract

This paper studies a rational inattention model in which a principal communicates with an uninformed agent about correlated multidimensional information. The principal agrees with the agent about appropriate actions to be taken given the states, though the players have potentially different preferences about relative importance of each dimension. I show that if the agent’s loss from suboptimal actions in one of the dimensions could be influenced by the principal, it would benefit the principal to increase the weight that agent places on the loss of this dimension when agent’s cost of processing information is either low or high, regardless of the principal’s weight on the loss. If cost is in the intermediate range, extra incentive provision on the dimension important to the agent could be harmful, particularly at the lowest cost such that decision quality of the other dimension would be affected. In this case, I characterize the condition under which it benefits the principal to provide (only) information about the dimension that is more important from her perspective but withhold information about the dimension more important to the agent. Withholding information of either dimension will not favor the principal when the agent’s cost is low or high. I find that the principal could be better off if the dimension important to the agent involves more complex information, but always gets a weakly lower payoff if the dimension important to the principal has more information. The model has implications for controlling incentives and information available in organizations.

Keywords: Information disclosure, Rational inattention, Costly communication.

JEL Codes: D82, D83, D91.

†Peicong (Keri) Hu: Department of Economics, University of California, San Diego, La Jolla, CA 92093, U.S.A. E-mail: p5hu@ucsd.edu. I am deeply indebted to Joel Sobel and Joel Watson for their continuing guidance and support. I am grateful to Snehal Banerjee and Jeremy Bertomeu for fruitful conversations and suggestions. I thank audiences at the UC San Diego Microeconomic theory seminar and Renee Bowen, Bradyn Breon-Drish, Simone Galperti, Ce Liu, Yibin Liu, Mario Milone, Korok Ray, Denis Shishkin, Ross Starr, Isabel Trevino, Emanuel Vespa, Yuehui Wang, and Chu Yu for helpful comments. All errors are my own.
1 Introduction

Information transmission is important in economic settings, but often impaired by difficulty in processing the information received. An agent who relies on a principal’s information to make decisions has to decide what issue to learn about if it is too costly to process all. Employees allocate their attention across activities in organizations. Jurors split their attention to a long list of evidence presented in trials. Students decide the part of lectures to which they will devote intensive attention. In many situations, the principal and the agent attach different values to decision quality of multiple issues. A tension arises in terms of what the agent should focus on. For instance, employees tend to devote excessive attention to information about short-term investment, though an informed decision about future opportunities is critical for the firm’s long-term growth and hence emphasized by the management team. This paper studies how the principal could motivate the agent to process information and improve his decision on the issue that is more important from the principal’s perspective.

In the model, the principal controls the agent’s access to information about a two-dimensional state of the world that is potentially correlated across dimensions. The agent takes an action on each dimension that would affect both players’ payoff. The agent is rationally inattentive, in the sense that he bears cost to process principal’s information but can choose the way in which he would acquire and use the information. In particular, agent allocates his attention between the two dimensions. The principal is “paternalistic” in the sense that she desires agent’s action on each dimension to be appropriate to the underlying state. So perfect information will be transmitted and both players will attain the highest payoff if agent is able to process all information provided by the principal. But if agent incurs attention cost and chooses to process only part of the information, a problem arises with regard to what information the agent would acquire. If agent agrees with the principal about (relative) importance of dimensions, the agent would pay sufficient attention to information that is more valuable from the principal’s perspective. If the players, however, have different stakes on each dimension, the agent would devote most of his attention to the dimension that he thinks to be important but ignore information that principal wants him to learn. Given the competing demands on attention, I examine potential strategies available for the principal to improve communication outcomes from her perspective.

Because the agent bears communication cost, the principal has limited ability to exert direct control over agent’s learning technology. It will be natural to consider increasing agent’s loss from misinterpreting principal’s signal to motivate learning. If the principal could raise the agent’s stakes on both dimensions, it is clear that the principal will benefit from more informed decisions by proportionally scaling up agent’s loss (from suboptimal actions) in each dimension. But if this is not feasible and principal could only influence agent’s incentive of learning about one of the dimensions, I show that strengthening agent’s incentive of learning on either one of
the dimensions would lead to a higher payoff for the principal when agent’s cost of processing information is either below a lower cutoff or above a higher cutoff. When the cost is sufficiently low, the agent processes information at a higher precision on the influenced dimension and at the same precision on the other dimension. So the decision quality on neither dimension will be undermined. When the cost is sufficiently high, the agent resists learning any information unless more incentive is provided. So a higher incentive surely benefits the principal, because at least some information is learned and incorporated into decision making. For medium costs, the result is not always clear though. If the principal is able to influence her valued dimension or if they agree about the relative importance of accurate decisions, the principal will get a strictly higher payoff by enhancing agent’s incentive. The effect, however, is indeterminate if principal could only influence the agent valued dimension. I characterize the condition under which the principal would get a strictly lower payoff by providing extra incentives. Shift in attention to the dimension important to the agent countervails more total effort being elicited and either force could dominate.

Apart from (directly) changing agent’s utility about decision quality, principal could influence agent’s learning process by controlling what will be disclosed. Excluding information about one dimension enhances the marginal value to learn about the other dimension if the two dimensions are correlated. I compare principal’s payoff when she discloses both dimensions and when she discloses only one of the dimensions. I show that it is never in principal’s interest to withhold information about her valued dimension. Doing so would eliminate the chance to learn principal valued information directly and dampen agent’s incentive to make efforts overall by restricting his choice to suboptimal strategies. Because agent has less incentive than the principal to compensate the loss by acquiring more information about the other dimension, principal will be strictly worse off if excluding her valued information. It, however, could benefit the principal to exclude agent valued information in some circumstances if and only if they value the decision quality of each dimension differently and agent’s cost is in the intermediate range. When cost is low, agent will devote sufficient attention to learn about both dimensions if information is fully disclosed. As cost increases, the gain from full disclosure due to more information available declines, because agent is not able to assimilate that much information even being provided. But the loss from full disclosure in terms of a less informed decision on principal valued dimension grows, because less information learned about the other dimension translates to less inferences implied for the principal valued dimension, adding another layer to a worse decision on this dimension. Finally, when cost is extremely high, the more effort induced by full disclosure becomes the dominant force, which favors no exclusion. Hence it is only beneficial to withhold information for an agent with medium cost.

The distinction of principal’s and agent’s values for (relative) decision quality has implications for the learning outcome when the information content, modeled by variance of the
state, changes. The principal’s payoff will be strictly decreasing in uncertainty about her valued dimension if the agent incurs medium cost and does learn something before the change. More effort induced from the agent and more attention allocated to principal valued dimension cannot offset the additional uncertainty. On the contrary, principal’s payoff could be strictly increasing in the uncertainty of the agent valued dimension. The agent is motivated to make greater efforts to learn about his valued dimension, which leads to more informed actions overall, and the increasing uncertainty about this dimension has less consequence on the principal’s payoff than the former case.

Different values attached to multiple issues are prevalent in economic settings. Myopic managers with horizon preferences conflicting with the firm’s collect information useful to boost current earnings but fail to incorporate information about long-term profitable investments. Lear and Abbott (2009) find that inefficiency in language programs arises from the fact that students and community partners weigh skills in distinct manners, though both of them desire a rewarding learning experience for all skills. Given the conflicting views about relative importance of issues, my theory examines strategies that can be adopted by the principal to attract and focus agent’s attention on what principal thinks to be important. I describe three real-world situations below which the theory potentially apply to.

The jury’s verdict is reached according to the weight of probative evidence. At common law it is occasionally permissible for the trial judge to express his or her opinion on the weight of the evidence, provided that the ultimate determination of facts is left to the jury. It is regarded as a valuable feature of the jury system that the judge (in some cases) provides the jury with advice and a review about the evidence from an analytical and dispassionate view (Walker, 1929). In this process, the judge might be able to influence the jury’s value for some issue. Nevertheless, the judge is not permitted to supply opinion for every aspect of the evidence, for example, prohibition of instructions about the weight given to the testimony of a witness (McDermott, 2017). But it is within the judge’s discretion whether to admit or exclude a questionable testimony to be used by jurors to arrive at a verdict (Bartol and Bartol, 2018).

In organizations, there is a long debate over the value and best use of the performance sensitive pays for managers in order to induce efforts (Stein, 1989; Holmström and Milgrom, 1991; Holmström, 1999; Murphy, 2013). A study by Koffarnus et al. (2013) is relevant to investment involves allocation of scarce resources, including the managers’ attention (Veldkamp, 2011; Kacperczyk et al., 2016). Edmans et al. (2013) attribute managers’ “short-termism” in this process to the horizon of managerial incentives. In response to managers’ incentive, the consultants are found to avoid advice that appears not to serve management’s interests and make recommendations that intend to benefit management rather than optimize benefits for the firm (Leiby, 2018). In contrast, I analyze strategies of a well-intentioned principal who paternalistically seeks to transmit information.

There is a large body of empirical literature about the effectiveness of limiting or disregarding instructions, though the theoretical foundation is limited.
the learning problem. Their result verifies that monetary incentives are effective in promoting engagement and achievement in a job-skills training program, which induce the participants to devote more attention. But a trade-off between “exploitation” and “exploration” activities has been recognized by research in organizational learning (March, 1991; Denrell and March, 2001; Fang et al., 2010). The skills of exploration or innovation is more difficult to be measured and rewarded than proficiency of using known solutions, at least in the short span of time. Hence the incentive for short-term rewards might trap employees in (identified) suboptimal solutions that cannot fit the changing business environment, because they then lack motivations to acquire skills for new solutions. Further, it is found that managers tend to bias learning towards exploitation particularly in the case of top-down knowledge inflows (Mom et al., 2007), which features the training scenarios. Thankfully, an emphasis on innovative activities (instead of reproduction) in training has been shown to enhance the potency, meaningfulness, impact, and autonomy of employees (Voegtlin et al., 2015; Sarri et al., 2010).

A similar problem has attracted vigorous attention of education policy makers and of standard-setting bodies of professional certification. The National Education Association urges schools to “take a fresh look at homework, and its potential for engaging students and improving student performance” (National Education Association, 2006). Researchers call upon teachers to “take into account grade-specific and developmental factors when determining the amount and kind of homework” (Cooper et al., 2006). Long homework assignments potentially strengthens students’ understanding about a subset of knowledge (Carroll, 1989), though evidence about the effectiveness is mixed (Dettmers et al., 2009). The evaluation should be based on the specific type of homework. In contrast to practice homework that helps review the content in lectures, extension homework embraces the idea of exploration activities. The extension homework, though potentially more complex (Lee Jr and Pruitt, 1979), aims at motivating students to transfer skills or concepts to a new problem (that has not been encountered before) (Cooper, 2015). Rosário et al. (2015) find that the extension homework positively impacted students’ mathematics achievement by leaving some problems figured out by students themselves, while the practice homework did not. Extension homework avoids students merely memorizing the formulae taught (instead of truly understanding the principles behind).

Hence a unified theory about effective communication is pending given the (controversial) use of judicial comments, exclusion of evidence, high-powered incentives, and a variety of homework assignments. This paper serves as a step to address management of attention in the presence of multiple issues. It shows that the following factors would matter to organize an

---

3Professional associations continue updating the content of examinations and the weight of sections to ensure that awardees would have sufficient knowledge about important issues. For instance, the AICPA is considering increasing the emphasis of SOCs (System Organization Controls) on the Audit and Business Environment and Concepts sections of the CPA Exam, so that newly licensed CPAs would be familiar with cybersecurity issues (Anders, 2020).
efficient information transmission process: the learner’s cost of understanding the information, the degree of alignment about relative value for different issues, the information content of each dimension, and the correlation across dimensions. I characterize conditions under which the practices above could be helpful to improve the learning outcome.

The remainder of the paper is organized as follows. Section 2 describes papers most closely related to my work. Section 3 illustrates the underlying logic through a simple mentor-employee example. Section 4 presents the basic model. Section 5 examines the effect of changing the agent’s learning incentive on one of the dimensions. Section 6 contains the analysis about restricting information about some dimension. Section 7 discusses comparative static results about the volatility of state. Section 8 contains concluding remarks. The appendix contains proofs that are not in the main text.

2 Related Literature

“Rational inattention” has been shown to be useful in explaining a variety of economic problems since the seminal work of Sims (2003) and has been applied to macroeconomics, finance, microeconomics, behavioral economics, labor economics, and political economy (Mackowiak et al., 2018). The multidimensional learning problem is recognized as one of the major problems that involve choice of attention. A profit-maximizing firm decides whether to attend to idiosyncratic shock or nominal shock when setting prices (Mackowiak and Wiederholt, 2009). An investor decides what assets to learn about when forming portfolios (Veldkamp, 2011). A voter decides what policy issue to focus on when making voting decisions (Matějka and Tabellini, 2017). A consumer decides what consumption opportunities to think about when allocating expenditures (Kőszegi and Matějka, 2020). Besides the information choice of the agent who has limited attention, my paper investigates how the trade-off in allocating attention between multiple issues would influence the disclosure strategy of an informed player.

To my knowledge, the first papers that study communication under rational inattention are Bloedel and Segal (2018) and Lipnowski et al. (2020a,b). Lipnowski et al. (2020a) show that a “benevolent” principal who has the same material preferences as the agent should optimally filter out information about the moderate state to enhance the marginal value of learning about extreme states for some parameter values in an example with three (linearly ordered) states. In the same three-state example, Lipnowski et al. (2020b) further characterize the optimal information policy that involves truthfully revealing the moderate state only when cost is small but revealing the extreme states only when cost grows to be larger. Bloedel and Segal (2018) find that if the principal and the agent have misaligned preferences, the principal would withhold information for two reasons: to manipulate the agent’s attention and to bias the agent’s decision. My paper considers a game with two-dimensional state of the
world and analyzes how conflict in relative importance across dimensions between the principal and the agent would affect principal’s strategy. Treating the decision problem about each dimension as one issue, I show that it would never benefit the principal to exclude either issue if they hold the same opinion about relative importance. So I highlight the role of preference weights about different issues in strategic communication. Moreover, the discussion about influencing agent’s loss from suboptimal actions, which serves as another tool to manage agent’s attention apart from withholding information, elaborates the trade-off in attention allocation between multiple (correlated) dimensions further. \cite{Lu2019} examines the trade-off in the financial reporting system between a noisy summary and precise details. The summary has less information content but requires less capacity to process. The tension that influences the principal’s optimal strategy is not different preferences about relative importance of multiple decisions.\footnote{In \cite{Lu2019}, the investor (agent) has a single decision that depends on the adaptation and coordination motives.} \cite{Bertomeu2020} study a voluntary disclosure game in which the investor optimally allocates fixed attention capacity. There is no moral hazard that is central to this model.

The literature about evidence exclusion provides insights about withholding information presented to the jury in trials. \cite{Lester2012} describe an example in which a piece of more informative evidence or a jury with superior ability could reduce the marginal benefit to seek other information, which leads to a strictly worse outcome for the society. In contrast, I show that it will always benefit the principal if communicating with an agent who has a lower cost to acquire information in this model, where the agent can \textit{flexibly} choose any learning strategy.\footnote{Instead of the need to read the whole piece of evidence, the agent could choose how much information he will process for each dimension in this model.} Given the flexible information choice, when they have exactly the same material preferences, it will harm the principal to exclude either dimension. \cite{Bull2019} study the merit of excluding \textit{misleading} hard evidence in the court when the litigant and fact-finder have beliefs that are out of alignment as another mechanism for withholding evidence to improve accuracy in the deliberations of a jury.

Multidimensional analysis in strategic decision making dates from the seminal paper of \cite{Holmstrom1991} that studies a principal-agent model (\textit{without communication}). A similar conclusion to this paper is that an increase in an agent’s compensation in any one task (\textit{enhancing incentive}) will cause some reallocation of attention away from other tasks. Instead of the increase in marginal cost of learning in \cite{Holmstrom1991}, what plays a role here is the reduction in marginal value of learning due to substitutability of information if belief about realization in the other dimension becomes more precise.\footnote{First, they assumed that dimensions are independent, i.e., the information about a particular dimension cannot help agent to infer about the state on the other dimension (so the \textit{direct} effect in my model is missing).}
(1991) assume that the measurement errors of tasks are independent but more effort in one task leads to a higher marginal cost in another task. Instead, the interesting case in this paper relates to the mutual influence of correlated dimensions. The entropy-based cost function that is not used in Holmström and Milgrom (1991) is tractable and appropriate for studying problems about information acquisition.

3 A Simple Example

In this section, I illustrate the logic of main insights using a simple example with binary states (and restricted information structure) which will be generalized in the full model. It will be clear from the example that players attain their highest possible payoff if the employee (agent) has full attention. If the employee, however, incurs cost to learn business knowledge and apply it to daily work, his decision will not be perfectly informed or appropriate to the underlying fundamental. It will benefit the mentor (principal) to offer higher powered incentives or to include more content in training in order to motivate learning when the employee weighs different knowledge in the same way as the mentor. The mentor, however, could induce a worse learning outcome from her perspective by providing extra incentive for making good decisions or communicating more information.

A mentor advises an employee about business knowledge that could either boost short-term earnings or uncover long-term profitable investments of the firm. Let $x_c \in \{0, 1\}$ be the return on current projects and $x_f \in \{0, 1\}$ be the return on potential future projects. Let $a_c \in [0, 1]$ be the scale of investment in the current projects and $a_f$ be the scale of investment in the future projects. The common prior about returns is $\Pr(x_i = 1) = 1/2$ for $i = c, f$. The correlation of the returns between current and future projects is given by $\Pr(x_i = t|x_j = t) = 2p^2$ for $t = 0, 1$ and $i = c, f$, where $p \in [1/2, \sqrt{2}/2]$. When $p = 1/2$, the returns of current and future projects are independent. When $p = \sqrt{2}/2$, the returns are perfectly correlated. Table 1 lists the prior

[In contrast to their analysis about interaction across dimensions, changing the incentive about either dimension will not affect the learning outcome of the other dimension if the correlation is zero in this paper, which is driven by the fact that the (optimally acquired) signal is independent across dimensions for the entropy based cost function.]

Second, Holmström and Milgrom (1991) assume that the agent’s costs depend only on the total effort or attention the agent devotes to all of his tasks (and the agent is not able to flexibly choose the information structure of signals). The reallocation of attention in my model results from substitutability of information about correlated dimensions, instead of the cross partial derivative of a strictly convex cost function. For future research, we could compare the models by departing from the assumption of cost being linear in entropy reduction and introducing convexity.

[The qualitative results extend beyond the binary precision levels – either perfectly informative or completely uninformative. The general model relaxes this specification and allows the employee to flexibly choose any information structure including a signal that is informative about two independent dimensions. The attention cost is a function of the amount of uncertainty reduction.]
belief about project returns. If the mentor commits to disclose (all) information about both current and future projects, the employee will choose probability \( e_i \in [0,1] \) to perfectly learn the return on project \( i \) for \( i = c, f \). That is, the employee will get a perfect signal about \( x_i \) with probability \( e_i \) and no signal about \( x_i \) with probability \( 1 - e_i \). After observing the signal (if any), the employee chooses the investment level \( a_i \) for \( i = c, f \).

Let \( u^M(x, a) = -(1/2)(a_c - x_c)^2 - (1/2)(a_f - x_f)^2 \) be the payoff of the mentor and \( u^E(x, a) = -\gamma_c(a_c - x_c)^2 - \gamma_f(a_f - x_f)^2 - c(e_c + e_f) \) be the payoff of the employee, where \( \gamma_i > 0 \ (i = c, f) \) is the weight/importance of project \( i \) in the employee’s utility and \( c \) is the unit cost. The linear-quadratic payoff function is the same as in the general model, except that specific weights are chosen for the mentor for simplicity. I assume that the mentor who is a senior of the firm places the same emphasis on short-term and long-term cash flows and hence has equal weight on the investment quality of either project. It is costly for the employee to acquire business knowledge and his cost is determined by the sum of efforts \( e_c + e_f \). The linear cost is the simplest functional form for us to analyze the employee’s multidimensional learning problem that assumes away the effect of increasing marginal cost or interdependence of cost across dimensions on effort choices.

Let us consider how the employee will allocate his attention of learning \( x_c \) and \( x_f \) when both signals are provided. The mentor and employee’s expected payoff as a function of employee’s efforts are given by

\[
\begin{align*}
\mathbb{E}u^M(e_c, e_f) &= -(1/2)[2p^2(1-2p^2)(1-e_c)e_f + (1/4)(1-e_c)(1-e_f)] \\
&\quad - (1/2)[2p^2(1-2p^2)e_c(1-e_f) + (1/4)(1-e_c)(1-e_f)] \\
\mathbb{E}u^E(e_c, e_f) &= -\gamma_c[2p^2(1-2p^2)(1-e_c)e_f + (1/4)(1-e_c)(1-e_f)] \\
&\quad - \gamma_f[2p^2(1-2p^2)e_c(1-e_f) + (1/4)(1-e_c)(1-e_f)] - c(e_c + e_f).
\end{align*}
\]

It is worth noting that the marginal value of devoting attention \( e_i \) for the employee is strictly decreasing in \( e_j \ (j \neq i) \) for \( p > 1/2 \), but does not depend on \( e_j \) for \( p = 1/2 \):

\[
\frac{\partial \mathbb{E}u^E(e_i, e_j)}{\partial e_i} = [2p^2(1-2p^2) - \frac{1}{4}](\gamma_i + \gamma_j)e_j + \frac{1}{4} - 2p^2(1-2p^2)]\gamma_j + \frac{1}{4} \gamma_i - c).
\]

I use the entropy-based cost function in the full model which is tractable in the linear-quadratic Gaussian environment for any (flexible) choice of the employee’s information structure. Further, given that cost depends (solely) on the amount of uncertainty reduction, the results will not tie to the sources of cost.

---

Table 1: The prior of returns

<table>
<thead>
<tr>
<th>Current project</th>
<th>Future project</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_c = 0 )</td>
<td>( x_f = 0 ) $\Pr(0,0) = p^2$</td>
</tr>
<tr>
<td>( x_c = 1 )</td>
<td>( \Pr(1,0) = 1/2-p^2 )</td>
</tr>
</tbody>
</table>

---

8I use the entropy-based cost function in the full model which is tractable in the linear-quadratic Gaussian environment for any (flexible) choice of the employee’s information structure. Further, given that cost depends (solely) on the amount of uncertainty reduction, the results will not tie to the sources of cost.
When returns $x_c$ and $x_f$ are independent, i.e., $p = 1/2$, the derivative with respect to $e_i$ is $(1/4)\gamma_i - c$ that does not depend on the attention level $e_j$. So the employee will determine the amount of attention devoted to each signal separately. When returns are correlated, i.e., $p > 1/2$, the coefficient $[2p^2(1 - 2p^2) - 1/4](\gamma_i + \gamma_j)$ is negative. In this case, more attention devoted to the project $j$ induces a more precise posterior belief about the return on project $i$, but makes it less valuable to learn the signal about $i$.

Assume without loss of generality that the current projects are (weakly) more important for the “myopic” employee, i.e., $\gamma_c \geq \gamma_f$. Note that there is no interior solution in this simple example, i.e., the employee’s effort choice $e_i$ is either 0 or 1 for $i = c, f$, because the marginal value of $e_i$ that is given by $\partial \mathbb{E}u^E(e_i, e_j)/\partial e_i$ is constant given $e_j$ for $j \neq i$. It then follows that if the employee devotes full attention to only one of the projects, it must be that $e_c = 1$ and $e_f = 0$ unless $\gamma_c = \gamma_f$. Furthermore, when $e_j = 1$, the derivative of $\mathbb{E}u^E(e_i, e_j)$ with respect to $e_i$ reduces to $2p^2(1 - 2p^2)\gamma_i - c$ which is the lowest value for $e_j \in [0, 1]$. The employee’s attention choice follows straightforwardly from these observations.

If $c \leq 2p^2(1 - 2p^2)\gamma_f$, the employee will choose full attention for both projects, i.e., $e_c = e_f = 1$. If $c > 2p^2(1 - 2p^2)\gamma_f$, full attention to the current projects ($e_c = 1$) would imply a negative marginal value of learning about future projects’ return. So the employee will completely ignore information about future projects, i.e., $e_f = 0$. When $2p^2(1 - 2p^2)\gamma_f < c \leq [1/4 - 2p^2(1 - 2p^2)]\gamma_f + (1/4)\gamma_c$, the employee will pay full attention to the return of current projects and no attention to that of future projects. When $c > [1/4 - 2p^2(1 - 2p^2)]\gamma_f + (1/4)\gamma_c$, the employee will learn nothing because of the high attention cost.

The employee’s information acquisition strategy is simple as described above. It illustrates that the marginal value of learning about something declines in the probability of getting informed about other correlated information, which has nothing to do with the attention cost. On one hand, the solution shares similarity with the general model that the employee will prioritize information generating a higher gain. If cost is sufficiently small, the employee will learn about both dimensions. If cost is somewhat larger, the employee will only learn the return about current projects in which he has a higher stake. On the other hand, the example is restrictive because the state of return is binary and the employee’s information structure is fixed. In other words, the employee is not able to attend to a signal that is composed of partial

---

9If there is some $e_i \in (0, 1)$ selected by the employee such that $\partial \mathbb{E}u^E(e_i, e_j)/\partial e_i = 0$ (given the value of $e_j$), then the employee’s payoff is constant regardless of the choice $e_i$. So the employee will get the same payoff by choosing $e_i = 1$ as well. If the original choice of $e_j$ gives the employee the highest payoff when $e_i = 1$, our restriction to the corner solution loses no generality and a higher effort favors the mentor. If another effort level $e_j'$ gives the employee a strictly higher payoff then $e_j$ when $e_i = 1$, the original strategy cannot be the optimal choice for the employee. Hence we can focus on the candidates 0 and 1 when characterizing the employee’s attention strategy.

10If $\gamma_c = \gamma_f$, the employee is indifferent between paying attention to information about either project.
information about returns of current and future projects, whereas he could devote half of the
time to learn business knowledge related to investment in current projects and half of the time
to learn information about future projects in real-world settings. In the math language, he is
able to pay attention to a signal about \((1/2)x_c + (1/2)x_f\) or any linear combination of \(x_c\) and \(x_f\). This flexibility of the employee’s attention strategy is captured by the general model in
which the loading of each dimension in the optimal signal is positive for any positive values of
\(\gamma_c\) and \(\gamma_f\).

Furthermore, the corner solution is driven by linearity of the attention cost, i.e., the marginal
cost of devoting more attention is not strictly increasing. In the general model, there is always an
interior solution due to decreasing returns to learning built into the entropy measure of cost. So
the attention level will not jump to zero immediately as cost increases. This has implications for
desirability of providing extra incentive to induce more attention from the mentor’s perspective.
In the current example, it always benefits the mentor to increase the employee’s weight of loss
on his preferred project \(\gamma_c\), because of a larger range of cost in which the employee will learn
some information. In contrast, negative consequences could be generated by a higher weight \(\gamma_c\)
for strictly convex cost. It is still true that the employee is motivated to acquire information
about \(x_c\) at some large cost while he declines to so for a lower \(\gamma_c\). But when the attention level
for future projects \(e_f\) has not reached zero, there are two countervailing effects of increase in
\(\gamma_c\). When the cost is not sufficiently small to induce \(e_f = 1\), the employee’s attention devoted
to future projects depends on the attention to current projects \(e_c\) and the (change in) weight
\(\gamma_c\). Let us take another look at the employee’s marginal utility of \(e_f\) (assuming the cost for
project \(i\) \((i = c, f)\) is instead \((ce_i^2)/2\):

\[
\frac{\partial \mathbb{E}u^E(e_c, e_f)}{\partial e_f} = \left[2p^2(1 - p^2) - \frac{1}{4}\right](\gamma_c + \gamma_f)e_c + \left[\frac{1}{4} - 2p^2(1 - 2p^2)\right]\gamma_c + \left(\frac{1}{4}\gamma_f - ce_f\right).
\]

The employee will devote more attention to current projects if \(\gamma_c\) increases, i.e., \(e_c\) increases
in \(\gamma_c\). For correlated returns, the first term above is negative. It becomes smaller for a higher
\(\gamma_c\) and \(e_c\). The second term is positive. It becomes larger for a higher \(\gamma_c\). So the overall
effect is indeterminate. I provide an example in Section 5 that the mentor could be worse off
by enhancing incentive to acquire information about current projects, which leads to a lower
marginal value for the employee to learn about future projects. But I show in the general model
that it will always benefit the mentor to increase the weight of future projects \(\gamma_f\) to which the
employee previously lacks attention.

Departing from the benchmark case in which the mentor discloses information about both
projects, we can also consider in this model whether a better learning outcome could be achieved
by withholding some content. Assume that before the vector \(\mathbf{x}\) is realized, the mentor announces
the training syllabus in terms of what information will be included. The employee could only
learn information about returns of projects within the “syllabus”. If only \(x_i\) is disclosed, the
employee’s expected payoff is given by

$$E u^E(e_i) = -\gamma_j [2p^2(1 - 2p^2)e_i + (1/4)(1 - e_i)] - \gamma_i (1/4)(1 - e_i) - ce_i.$$  

The marginal value of learning about $e_i$ is

$$\frac{\partial E u^E(e_i)}{\partial e_i} = \left[ \frac{1}{4} - 2p^2(1 - 2p^2) \right] \gamma_j + \left( \frac{1}{4} \gamma_i - c \right).$$

which is the employee’s marginal utility when $e_j = 0$. Recall that the marginal value of devoting attention $e_i$ is strictly decreasing in $e_j$ ($j \neq i$) for correlated returns. So it is clear that excluding dimension $j$ would boost learning about $i$. The mentor then faces a trade-off: withholding information about current projects (that the employee biases towards) will enhance his incentive to learn about investment opportunities in the future, but will make less information available to be potentially used in decisions and reduces the employee’s overall effort. Section 6 deals with the mentor’s optimal strategy under this trade-off.

This example suggests the intuition why awarding more bonus for current projects or disclosing information that the employee has a higher stake (than the mentor) could lead to unintended consequences. I will develop the full model in the next section and provide formal results about how the mentor should direct the employee’s attention after introducing the model. I will use “principal” and “agent” in the general model who are mentor and employee, respectively, in this example.

### 4 The Benchmark Model

This section describes the disclosure model in which information and actions are two-dimensional and the agent processes information at some cost. Proposition 1 characterizes agent’s information acquisition strategy for the linear-quadratic utility function with weighted dimensions. The characterization is simple and deals with potentially correlated realizations across dimensions.  

#### 4.1 The Preliminaries

There are two players, principal and agent. The principal is potentially informed of the realization of a two-dimensional state of the world $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$. Assume that the prior uncertainty about $\boldsymbol{x}$ is multivariate Gaussian, where $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \Psi)$ and the variance-covariance

\footnote{Kőszegi and Matějka (2020) focuses on the allocation of attention driven by preferences only and assumes that the dimensions are independent (and the variance of each dimension is the same). Instead, this paper analyzes how the correlation across dimensions affects agent’s learning process and the principal’s choice of information disclosure and assumes that there is no interaction term of the losses across dimensions in the players’ utility function, i.e., the losses from misperception in the two dimensions are separable. The characterization here is particularly simple and intuitive in this environment.}
matrix is given by $\Psi = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$ with $-1 \leq \rho \leq 1$. We are able to derive full analytical solutions for Gaussian uncertainty.

We leave detailed discussions about the potential strategy available for the principal to Sections 5 and 6 but focus on information acquisition of the agent in this section assuming that perfect information about both dimensions is disclosed by the principal. The game (starting with the agent’s choice) proceeds as follows. The agent chooses an information structure about the joint distribution of the state $x$ and the signal he observes. Then the state is realized and disclosed to the agent. The agent observes the signal realization given by his information structure and chooses a vector of actions $a = (a_1, a_2) \in \mathbb{R}^2$. Finally, their payoffs are realized.

For Gaussian priors, Sims (2003) shows that the optimal signal for an agent with a linear-quadratic utility function is Gaussian as well among all possible information structures. Furthermore, the posterior uncertainty generated by the optimal signaling structure is Gaussian and has a constant, i.e., state-independent, variance-covariance matrix. Since we assume that the principal discloses perfect information, the agent’s posterior belief generated by the optimal signal is Gaussian in this environment. Hence, without loss of generality, we could simplify the agent’s problem by the following reduced form of the attention cost.

The choice variable of the agent is the posterior variance-covariance matrix $\Sigma$ subject to the no-forgetting constraint that $\Psi - \Sigma$ is positive semi-definite. The no-forgetting constraint requires that the posterior has to be (weakly) more precise than the prior. Let the posterior distribution of the state $x$ be $N(\tilde{x}, \Sigma)$, where $\tilde{x}$ is mean of the posterior belief and the variance-covariance matrix $\Sigma = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\rho} \hat{\sigma}_1 \hat{\sigma}_2 \\ \hat{\rho} \hat{\sigma}_1 \hat{\sigma}_2 & \hat{\sigma}_2^2 \end{pmatrix}$ with $-1 \leq \hat{\rho} \leq 1$. Denoting by $|\cdot|$ the determinant of a matrix, the cost of attention for the multivariate Gaussian distribution is given by

$$(c/2)(\log(2\pi e)^2|\psi| - \log(2\pi e)^2|\Sigma|) = (c/2)(\log|\psi| - \log|\Sigma|),$$

where $c \geq 0$ is the agent’s unit attention cost for reduction in the entropy of his belief.

The principal’s payoff is given by

$$u^P(a, x) = -\left[\gamma_1^P (a_1 - x_1)^2 + \gamma_2^P (a_2 - x_2)^2\right],$$

where $\gamma_1^P, \gamma_2^P > 0$ parameterize the principal’s valuation for the two dimensions; and the agent’s

\[\text{principal’s loss from suboptimal action}\]

\[\text{agent’s loss from suboptimal action}\]

\[\text{The literature has well established that “choosing the signal variances is the same as choosing a variance of the posterior belief” (Veldkamp, 2011). Further, one rule for Bayesian updating with normal variables is that the precision of the posterior belief is the sum of the precisions of prior and signals.}\]

\[\text{I adopted the simplest (linear) entropy cost function to illustrate the intuitions. The results are robust with respect to a broader range of alternatives, including the convex entropy-based cost function.}\]
payoff is given by

\[ u^A(a, x, \Sigma) = -\left[ \gamma_1^A (a_1 - x_1)^2 + \gamma_2^A (a_2 - x_2)^2 \right] \]

agent’s loss from suboptimal action

\[- \left( \frac{c}{2} \right) (\log |\Psi| - \log |\Sigma|), \]

agent’s information cost

where \( \gamma_1^A, \gamma_2^A > 0 \) parameterize the agent’s valuation for the two dimensions. Both players want agent’s actions to be matched to the state realizations, but they have potentially different stakes on each dimension. Without loss of generality, assume that the principal weakly prefers accurate decisions on dimension one relative to those on dimension two more than the agent. In a relative sense, dimension one is called the principal preferred dimension and dimension two is called the agent preferred dimension.

**Assumption 1.** Assume that \( \gamma_1^P / \gamma_2^P \geq \gamma_1^A / \gamma_2^A \).

It follows that \( \gamma_1^P / \gamma_1^A \geq \gamma_2^P / \gamma_2^A \), which will be useful throughout the analysis about the principal’s problem.

The solution concept is perfect Bayesian equilibrium.

### 4.2 Agent’s information acquisition strategy

The solution to the agent’s problem is based on the water-filling algorithm used in information theory (see Telatar (1999); Cover and Thomas (2006)). If the attention cost is extremely high, the agent will not process any information. If the cost is somewhat lower, the agent will choose to process information about one dimension only that is a linear combination (“principal component”) of the state vector. If the cost is even lower, the agent will process information about two independent linear combinations of the state.\(^{14}\)

The agent maximizes his expected utility \( u^A(a, x, \Sigma) \) that depends on the vector of states \( x \), the vector of actions \( a \), and his information cost. Let us solve the agent’s problem backwards assuming that the principal discloses both dimensions.

Given a posterior belief about the state realization \( x \sim \mathcal{N}(\tilde{x}, \Sigma) \), the agent would choose a vector of actions \( a \) equal to the posterior mean \( \tilde{x} \) in order to minimize the quadratic loss function.

**Lemma 1.** Given a posterior belief about the state \( x \sim \mathcal{N}(\tilde{x}, \Sigma) \), the agent’s action \( a_i \) is equal to the posterior mean \( \tilde{x}_i \) for \( i = 1, 2 \).

Because the expectation of \( (\tilde{x}_i - x_i)^2 \) is the agent’s posterior variance, we can express the payoff function of the agent (and the principal) in terms of his posterior variance. Suppressing\(^{14}\)

---

\(^{14}\)The application to different problems that has similar spirit can be found in Kacperczyk et al. (2016), Miao et al. (2019), and Kőszegi and Matějka (2020).
the constant term of the prior entropy, the agent’s problem is
\[
\max_\Sigma -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{c}{2} \log |\Sigma| \\
\text{subject to } \Psi - \Sigma \text{ is positive semi-definite.}
\]

(1)

We transform the coordinates to another basis so that the objective (including the cost term) is shown to be additively separable in the posterior variance of each dimension.

Given that the losses from misperceptions (and suboptimal actions) are weighted by \(\gamma_1^A\) and \(\gamma_2^A\), we need to incorporate the weights by which imprecision of decisions is translated into agent’s losses, in order to find out the directions in which losses are the largest (principal components). To capture this effect, let \(w = (w_1, w_2) = (\sqrt{\gamma_1^A} x_1, \sqrt{\gamma_2^A} x_2)\). Then the posterior variance-covariance matrix of the vector \(w\) is given by
\[
\Phi = \begin{pmatrix}
\frac{\gamma_1^A}{\gamma_1^A} \sigma_1^2 & \sqrt{\frac{\gamma_1^A}{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} } \\
\sqrt{\frac{\gamma_1^A}{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2}} & \frac{\gamma_2^A}{\gamma_2^A} \sigma_2^2
\end{pmatrix}.
\]

Let \(v^i\) be the orthonormal basis of eigenvectors of \(\Phi\) with the eigenvalue corresponding to \(v^i\) denoted by \(\Lambda_i\) \((i = 1, 2)\). Assume that \(\Lambda_1 \geq \Lambda_2\). So the eigenvector \(v^1\) \((v^2)\) gives the loadings of \(w_1\) and \(w_2\) on the first \(\text{(second)}\) principal component that is the direction in space along which projections have the larger \(\text{(smaller)}\) prior variance. Let \(U\) be the \((\text{unitary})\) matrix that consists of the eigenvectors of \(\Phi\), i.e., \(U = (v^1 \ v^2)\), and \(\Lambda\) be the \((\text{diagonal})\) matrix such that \(\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}\). I claim that the agent’s optimal information strategy is to acquire \((\text{independent})\) signals of \(v^i \cdot w\) such that the posterior variance of \(v^i \cdot w\) is \(\min(c/2, \Lambda_i)\).

**Proposition 1.** The optimal information strategy is to acquire independent signals of \(v^i \cdot w\) such that the posterior variance of \(v^i \cdot w\) is \(\min(c/2, \Lambda_i)\), where \(v^i\) is the orthonormal basis of eigenvectors of \(\Phi\) with the eigenvalue corresponding to \(v^i\) denoted by \(\Lambda_i\), and \(w = (\sqrt{\gamma_1^A} x_1, \sqrt{\gamma_2^A} x_2)\).

This signal is optimal even if the agent is able to choose any signal and the eigenvectors obtained by eigendecomposition are not restricted to be the same for prior and posterior variance-covariance matrices. Clearly, the two vectors \(v^1 \cdot w\) and \(v^2 \cdot w\) are independent, which is evident from the definition that \(U \Lambda U' = \Phi\) and \(U\) being idempotent by construction. It is nevertheless possible that for two independent dimensions \((\text{principal components})\), the optimal signal of each dimension contains information about the other dimension, as pointed out by Van Nieuwerburgh and Veldkamp (2010) and Köszegi and Mátéjka (2020). I show for this game that \(v^1\) and \(v^2\) are actually the eigenvectors of posterior variance-covariance matrix of the vector \(w\) generated by the optimal signal as well. In other words, the agent acquires information about \(v^1 \cdot w\) and \(v^2 \cdot w\) separately, subject to the no-forgetting constraint that the posterior variance cannot exceed the prior variance \(\Lambda_i\) of the principal component \(i\). Intuitively, the
agent should prioritize acquisition of information that would matter the most, i.e., generating the largest loss if being misperceived. Loadings given by vectors \( v^1 \) and \( v^2 \) capture the agent’s preferences about the weighted signal and prior correlation between the two dimensions.

By Proposition 1 when the cost is high enough such that \( c/2 > \Lambda_1 \), the agent will not process any information. When half of the parameter \( c \) is between \( \Lambda_2 \) and \( \Lambda_1 \), the agent will only learn information about the first principal component that has more prior uncertainty and the amount of information acquired is \( \Lambda_1 - c/2 \). The dimension with more prior uncertainty will be more valuable for the agent to learn about due to diminishing returns to learning modeled by the entropy technology. When the cost is low enough such that \( c/2 \leq \Lambda_2 \), the resulting posterior variance \( c/2 \) of both principal components would be smaller than the prior variance \( \Lambda_i \) \((i = 1, 2)\). But the agent still learns more information about the first principal component than about the second principal component, i.e., \( \Lambda_1 - c/2 \geq \Lambda_2 - c/2 \).

I close this section with calculating the principal’s expected payoff given that agent learns optimally. Let the posterior variance-covariance matrix of \( w \) be \( \Xi = \begin{pmatrix} \gamma_1^A \sigma_1^2 & \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \\ \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} & \gamma_2^A \sigma_2^2 \end{pmatrix} \).

Recall that loadings of \( w_1 \) and \( w_2 \) on the principal components are given by the eigenvectors \( v_1 \) and \( v_2 \). Let \( v_i^j \) be the weight of \( w_i \) on the \( j^{th} \) principal component. So the first principal component is given by \( v_1^1 w_1 + v_1^2 w_2 \) and the second principal component is given by \( v_2^1 w_1 + v_2^2 w_2 \). We can derive the posterior variance \( \Xi \) for the underlying (correlated) states by Proposition 1:

\[
\Xi = U \begin{pmatrix} \min(c/2, \Lambda_1) & 0 \\ 0 & \min(c/2, \Lambda_2) \end{pmatrix} U' = \begin{pmatrix} \min(c/2, \Lambda_1)(v_1^1)^2 + \min(c/2, \Lambda_2)(v_1^2)^2 & \min(c/2, \Lambda_1)v_1^1 v_1^2 + \min(c/2, \Lambda_2)v_1^2 v_2^2 \\ \min(c/2, \Lambda_1)v_1^1 v_2^1 + \min(c/2, \Lambda_2)v_1^2 v_2^2 & \min(c/2, \Lambda_1)(v_2^1)^2 + \min(c/2, \Lambda_2)(v_2^2)^2 \end{pmatrix}.
\]

(2)

The following lemma will be useful throughout the analysis.

**Lemma 2.** The prior variances of the principal components are

\[
\Lambda_1 = \frac{1}{2} \left( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \right);
\]

\[
\Lambda_2 = \frac{1}{2} \left( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \right).
\]

Furthermore, the eigenvectors satisfy \( v_1^2 = v_2^1 \) and \( v_2^2 = -v_1^1 \).

I state the principal’s expected payoff given that the agent chooses the optimal information strategy in Proposition 1.
Lemma 3. The principal’s expected payoff with full disclosure is given by

\[
\mathbb{E}u^P = \begin{cases} 
-\frac{c}{2}(\frac{\gamma_1^P}{\gamma_1} + \frac{\gamma_2^P}{\gamma_2}) & \text{if } c/2 \leq \Lambda_2 \\
-(\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) + (\Lambda_1 - \frac{c}{2})[(\frac{\gamma_1^P}{\gamma_1^2})(v_1^1)^2 + \frac{\gamma_2^P}{\gamma_2^2}(1 - (v_1^1)^2)] & \text{if } \Lambda_2 < c/2 \leq \Lambda_1 \\
-(\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) & \text{if } c/2 > \Lambda_1 
\end{cases}
\]

If \(c/2\) is less than or equal to the smaller eigenvalue \(\Lambda_2\) which is the prior variance of the second principal component, the posterior variances of both principal components reduce to \(c/2\). Then the posterior variances of the underlying states are \(c/(2\gamma_1^A)\) and \(c/(2\gamma_2^A)\), respectively. So the principal’s expected payoff weighted by her own value is given by \(-[(\gamma_1^P c)/(2\gamma_1^A) + (\gamma_2^P c)/(2\gamma_2^A)]\). If \(c/2\) is between \(\Lambda_2\) and \(\Lambda_1\), the agent only processes information about the first principal component. The first term \(-(\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2)\) is principal’s expected payoff without any information being transmitted. The second term is the gain from reduction in uncertainty about the first principal component. If \(c/2\) is greater than \(\Lambda_1\), it is not worthwhile for the agent to acquire any information, because the prior variances have been lower than the target variance \(c/2\). Figure 1 illustrates the agent’s learning strategy.

Figure 1: \(c\) versus \(u^P\)

In Sections 5 and 6 I discuss two potential tools for the principal to “improve” the agent’s learning outcome when there are limited transfers available. The conflict arises from agent’s cost to process information and disagreement over the relative value of information for them. I show that if principal is able to influence the agent’s learning incentive on one dimension, it would always benefit the principal to increase agent’s (weighted) loss on this dimension regardless of whether it is the principal preferred one, when the cost parameter is either low or high. When the cost is in the intermediate range, the other tool could be useful, that is, to
withhold information about the agent preferred dimension. The following sections discuss each case separately.

5 Influencing Agent’s Payoff

In this section, the principal is assumed to have an ability to influence the agent’s weight of loss. The question for the principal lies in the value of exerting influences and how the principal should utilize it to motivate the agent to acquire (principal valued) information.

Presumably, if the principal could control agent’s weights on both dimensions $\gamma_A^1$ and $\gamma_A^2$, it would lead to a weakly higher payoff for the principal than if the incentive on only one of the dimensions could be influenced. Furthermore, it is clearly beneficial to the principal to proportionally scale up the agent’s loss on both dimensions. The problem arises when agent’s incentive on only one of the dimensions could be influenced, due to institutional reasons or feasibility constraints. In the introduction, I described this problem in the contexts of judiciary law, management incentives, and teaching policies. For example, it could be prohibited to influence the weight assigned to some evidence or hard to reward certain skills. This section offers an answer about the principal’s proper choice if she could influence one dimension only.

Most of the results in this section are local in the sense that my focus will be the effect of changing agent’s incentive on the margin. On one hand, it will be certainly in principal’s interest if she could bring agent’s loss to extremely high when the two dimensions are sufficiently correlated. The agent would have an incentive to learn precisely the impacted dimension and infer realization of the other dimension as well. On the other hand, it is often unrealistic that the principal has an ability to arbitrarily influence agent’s preferences. The judge is prohibited to determine the credibility of each witness and the weight given to a testimony on behalf of the jury. The teacher cannot exclusively influence students’ preferences towards any skills. Hence it seems to be more valuable for us to consider the (local) effect exerted upon agent’s pre-given payoff.

Suppose that the principal could marginally change agent’s weight of loss on a dimension $i$ before state realization. The dimension which is controllable by the principal is exogenously fixed. The agent then acquires information based on his new preferences. The question for the principal is how her payoff will change in the parameter $\gamma_i^r$ that represents agent’s value for information about dimension $i$. Proposition 2 shows that if the two dimensions are perfectly correlated or have no correlation at all, it strictly benefits the principal to enhance agent’s incentive of learning about dimension $i$, whichever the controllable dimension $i$ is.

**Proposition 2.** If $\rho = \pm 1$, agent’s posterior variances on both dimensions are decreasing in $\gamma_i^r$ for $i = 1, 2$. If $\rho = 0$, agent’s posterior variance on dimension $i$ is decreasing in $\gamma_i^r$ but the

\[\text{Please refer to \textit{Vermont General Jury Instructions}}\]
posterior variance on dimension $j \neq i$ is constant, for $i = 1, 2$. In both cases, principal’s payoff is increasing in $\gamma_i^r$ for $i = 1, 2$.

When the information is perfectly correlated, the principal is essentially raising agent’s incentive to learn both dimensions. When the information is completely uncorrelated, agent learns more about the controlled dimension $i$ but acquire the same amount of information on the uncontrolled dimension. As a result, the principal would gain a higher total payoff.

The results will not be as straightforward when states of the two dimensions are imperfectly correlated. Proposition 3 shows that the result is definite when principal is able to influence the dimension that she values more in the relative sense, i.e., $i = 1$.

**Proposition 3.** Fix agent’s value parameter on dimension two $\gamma_2^A$. When $c/2 \leq \Lambda_1$, the principal’s payoff is strictly increasing in $\gamma_1^A$. When $c/2 > \Lambda_1$, agent does not learn any information until $\Lambda_1$ reaches the threshold $c/2$ as $\gamma_1^A$ increases. The principal gains a higher payoff (than if no information is acquired) from here onwards.

There are two consequences on agent’s learning behavior when $\gamma_1^A$ increases. On one hand, agent will be motivated to devote more total effort to internalizing principal’s information. As a result, the decisions on both dimensions will be more informed. On the other hand, agent will be induced to reallocate some attention to the increasingly important dimension one, which is favored by the principal. The two effects add together, leading to a higher payoff for the principal.

Mathematically, when $c/2 \leq \Lambda_2$, agent acquires information about both principal components. As $\gamma_1^A$ increases, the loading $v_1^1$ of dimension one on the first principal component will increase, which implies more (less) information about dimension one (two) is acquired. However, the loading $v_2^2$ of dimension two on the second principal component will increase as well, since $v_2^2$ is equal to $v_1^1$. So along the direction of the second principal component, agent learns more (less) about dimension two (one). Overall, the same amount of information will be acquired on dimension two as before. But agent will learn the state of dimension one more precisely, because magnitude of the change in prior variances of principal components differs.

It is worth noting that because prior variances $\Lambda_1$ and $\Lambda_2$ of the principal components are strictly increasing in $\gamma_1^A$, $c/2$ must be always less than $\Lambda_2$ and principal’s payoff will then be monotonically increasing in $\gamma_1^A$.

When $\Lambda_2 < c/2 \leq \Lambda_1$, the agent will only acquire information about the first principal component. In this case, the competing demands on attention would induce agent to acquire more information about dimension one at the expense of a potentially less informed decision on dimension two.\footnote{The net effect on dimension two depends on whether more inferences driven by the increase in total effort would outweigh the substitution of attention.}

Because principal values dimension one relative to dimension two more
than the agent, this shift in attention will be beneficial to the principal. Furthermore, the agent will be induced to devote more overall attention. The prior variance $\Lambda_1$ of the first principal component becomes larger, as incorporates agent’s weight of loss. Therefore, the principal will gain a strictly higher payoff in this scenario.

When $c/2 > \Lambda_1$, the agent does not find valuable to acquire any information. The value of information in terms of more informed decisions is outweighed by the large cost that agent has to bear. So agent will not learn anything initially. Nevertheless, as $\gamma_1^A$ increases, the loss for the agent from suboptimal actions becomes higher. Finally, agent will start to pay attention to principal’s disclosure when $c/2$ exceeds $\Lambda_1$. In this sense, raising the weight $\gamma_1^A$ favors the principal as well, otherwise no information will ever be learned and used in decision making.

The effect of increasing agent’s loss from misperception, however, is not clear if the dimension on which the payoff could be influenced is the agent valued one, i.e., dimension two. Given that agent values dimension two (relative to dimension one) more than the principal, reallocation of attention would act against the benefit of increase in effort elicited from the principal’s point of view. Shift in attention away from dimension one and the resulting worse decision on this dimension could then lead to a worse outcome for the principal. I first show in Proposition 4 that principal’s payoff is increasing in agent’s weight $\gamma_2^A$ on dimension two if they value the two dimensions in the same (relative) way. An indeterminate result follows in Proposition 5 where principal is assumed to value dimension one relative to dimension two strictly more than the agent. A numerical result illustrates that the sign of derivative could change in the undetermined case.

**Proposition 4.** If $\gamma_1^P/\gamma_2^P = \gamma_1^A/\gamma_2^A$, then principal’s payoff is increasing in $\gamma_2^A$.

The result is intuitive because reallocation of attention affects the principal and the agent in the same way. The agent’s optimal choice of attention will then be the principal preferred decision as well. But if principal values dimension one more than agent, the result is not clear for intermediate values of cost.

**Proposition 5.** Fix agent’s value parameter on dimension one $\gamma_1^A$. Suppose that $0 < |\rho| < 1$ and $\gamma_1^P/\gamma_2^P > \gamma_1^A/\gamma_2^A$.

(i) When $c/2 \leq \Lambda_2$, agent’s posterior variance on dimension two is strictly decreasing in $\gamma_2^A$, while the posterior variance on dimension one is constant. The principal’s payoff is strictly increasing in $\gamma_2^A$.

(ii) When $\Lambda_2 < c/2 \leq \Lambda_1$, the effect of change in $\gamma_2^A$ on principal’s payoff is indeterminate in general. The principal’s payoff can increase or decrease in $\gamma_2^A$.

(iii) When $c/2 > \Lambda_1$, agent does not learn any information until $\Lambda_1$ reaches the threshold $c/2$ as $\gamma_2^A$ increases. The principal gains a higher payoff (than if no information is acquired) from here onwards.
The arguments for (i) and (iii) are similar to Proposition 3 except that the roles of the two dimensions are swapped. But the effect of influencing agent’s incentive now becomes unclear in (ii), because the two forces, the additional attention elicited by principal’s influence and the variation in fraction allocated to each dimension, countervail each other from the principal’s perspective. The intuition is elaborated below.

Increasing agent’s incentive to learn dimension two has two opposing effects on the decision quality of dimension one. On one hand, agent could directly infer more information about dimension one from the more precise observation about dimension two. On the other hand, a higher precision on dimension two would increase the posterior precision on dimension one if the states are correlated, lowering the attention required to achieve the same posterior precision. Hence learning more about dimension two indirectly dampens agent’s incentive to learn about dimension one given substitutability of the information. Even if the loss from misperception on each dimension is additively separable, the marginal value of learning about dimension one declines by diminishing marginal return to learning.

Let us further decompose the impact on principal’s payoff as follows:

\[
\frac{\partial u^P}{\partial \gamma_2} = \left( v_1 \right)^2 \gamma_1^P \frac{\partial \Lambda_1}{\partial \gamma_2} + \left( 1 - \left( v_1 \right)^2 \right) \frac{\gamma_2^P \partial \Lambda_1}{\gamma_2^P \partial \gamma_2} \frac{\partial (v_1^2)}{\partial \gamma_2} + \left( \Lambda_1 - \frac{c}{2} \right) \frac{\gamma_1^P \gamma_2^A}{\gamma_1^A \gamma_2^A} \frac{\partial (v_1^2)}{\partial \gamma_2^A}.
\]

Better decision on dimension one (fixing loadings) + Change in decision quality on dimension two (fixing loadings) + Change in principal’s payoff from more overall attention if fixing loadings of each dimension

Decrease in principal’s payoff due to reallocation of attention

The prior variance of principal components is increasing in \( \gamma_2^A \), i.e., \( \partial \Lambda_1 / \partial \gamma_2^A > 0 \) which implies that more effort is expensed. Recall that the fraction of information learned by the agent on each dimension is given by the respective loading, \( v_1^2 \) and \( v_2^2 \). So the decision quality on dimension one improves if we fix the loadings of each dimension, which is given by the first term. The sign of the second term is indeterminate, as the change in loadings is not taken into account. I claim that the third term that captures agent substituting attention away from dimension one is negative. This represents principal’s loss from agent reallocating scarce attention, as principal values dimension one more than the agent. Giving a look at the sign of each component, it is clear that \( \Lambda_1 - \frac{c}{2} \) is positive for intermediate costs, \( \partial (v_1^2) / \partial \gamma_2^A \) is negative and \( \gamma_1^P / \gamma_1^A \) is greater than \( \gamma_2^P / \gamma_2^A \) by Assumption 1. Hence the overall impact on principal’s payoff would depend on parameters.

Figure 2 provides an example in which principal’s payoff first increases and then decreases in \( \gamma_2^A \) within the range of \( \Lambda_2 < \frac{c}{2} \leq \Lambda_1 \). Corollary 1 shows that when \( \frac{c}{2} \) is close to \( \Lambda_1 \),

---

17 Please see Lemma 7 in the Appendix for details.
18 Please see Lemma 8 in the Appendix for details.
principal’s payoff will be strictly increasing in $\gamma_2^A$, as the effect of total effort would dominate. The result is proved by observing that principal’s marginal utility in $\gamma_2^A$ is strictly increasing in $c$.[19]

![Incentive on dimension two versus Principal payoff](image)

**Figure 2:** $\gamma_2^A$ versus $u^P$

Corollary 1. There is some $c' \in (\Lambda_2, \Lambda_1)$ such that principal’s payoff is (locally) strictly increasing in $\gamma_2^A$ for any $c \in (c', \Lambda_1)$.

Hence when cost is sufficiently large, more attention being elicited by the higher incentive would dominate the substitution effect, which favors the principal. But for a lower cost, which force would dominate depends on parameters.

6  Restriction on Disclosure

Apart from (directly) influencing agent’s value about learning information, it is natural for the principal to also consider what information to be disclosed. Because they attach values to each dimension differently, the agent may focus his attention on something that is not the emphasis of the principal. A straightforward solution available for the principal is to withhold disclosure of some information, hoping to refocus agent’s attention (on principal valued issues). Judges withhold some evidence from jurors in trials; supervisors organize activities for front-line staff to master new skills which will be used in the future; teachers leave some room for students to apply knowledge to novel situations. This section is intended to address which dimension

---

[19]If $c/2$ is close to $\Lambda_2$, the effect of substitution could dominate.
could be ever optimally excluded and characterize conditions under which it is sensible for the principal to withhold information.

To analyze principal’s disclosure strategy, I assume that she chooses whether to disclose perfect information or no information on each dimension before any state realization. Upon observing the principal’s choice, the agent decides his attention strategy (that is a joint distribution between the signal he observes and the true signal). When the principal withholds information about one dimension but discloses perfect information about the other dimension, the resulting posterior beliefs would still be Gaussian if agent optimally chooses the signal. This will be clear from the decomposition below that transforms the two dimensional problem into one with a random loss from suboptimal actions on the disclosed dimension plus a constant loss from another independent dimension. So the theorem in Sims (2003) applies here as well, implying that it is without loss of generality to restrict our attention to a Gaussian signal (on the disclosed dimension). Hence the agent is still choosing a posterior variance-covariance matrix subject to a no-forgetting constraint.

6.1 Agent’s strategy if one dimension is excluded

Suppose that the principal only discloses the state on dimension $i$, but withholds any information on dimension $j$. For the state $x = \begin{pmatrix} x_i \\ x_j \end{pmatrix}$ such that $x \sim \mathcal{N}(0, \Psi)$ where $\Psi = \begin{pmatrix} \sigma_i^2 & \rho \sigma_i \sigma_j \\ \rho \sigma_i \sigma_j & \sigma_j^2 \end{pmatrix}$, we can write the vector in terms of two independent random variables $z_i$ and $z_j$:

$$x_i = \sigma_i z_i \quad \text{and} \quad x_j = \sigma_j (\rho z_i + \sqrt{1 - \rho^2} z_j).$$

In the matrix form,

$$\begin{pmatrix} x_i \\ x_j \end{pmatrix} = \begin{pmatrix} \sigma_i & 0 \\ \sigma_j \rho & \sigma_j \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} z_i \\ z_j \end{pmatrix}. $$

The prior variance-covariance matrix of $z = \begin{pmatrix} z_i \\ z_j \end{pmatrix}$ is the identity matrix $I$. The agent’s problem could be solved in a simple way with $z$ being the argument, as the principal’s signal does not involve anything about $z_j$.

**Lemma 4.** The optimal information strategy if only dimension $i$ is disclosed is to acquire the signal of $i$ such that the posterior variance on dimension $i$ is $\sigma_i^2 \min(1, \frac{c}{2(\sigma_i^2 \gamma_i + \sigma_j^2 \rho^2 \gamma_j)}).$

The agent’s loss could be written as a quadratic function of a Gaussian random variable $z_i$. So the optimal signal for the agent is still Gaussian. Lemma 4 shows how agent chooses the posterior variance of $x_i$ to trade off his loss from misperception against the information cost.
The principal’s payoff of disclosing only dimension $i$ is then given by
\[-(\sigma_i^2 \gamma_i^s + \sigma_j^2 \rho^2 \gamma_j^s) \min(1, \frac{c}{2(\sigma_i^2 \gamma_i^r + \sigma_j^2 \rho^2 \gamma_j^r)}) - \sigma_j^2 (1 - \rho^2) \gamma_j^s.\]

We can calculate principal’s expected payoff if she discloses dimension one only or dimension two only. The comparison with full disclosure is discussed in the next section.

6.2 Principal’s disclosure strategy

If states of the two dimensions are perfectly correlated, the agent could “precisely” infer the realization of one dimension if observing that of the other dimension. Essentially, agent only needs access to the information about one of the dimensions only. If states are independent across dimensions, agent processes information about the two dimensions separately. So if the principal excludes the signal of some dimension, the agent will not be able to learn anything about this dimension but still learn the same amount of information about the other dimension. Hence it is not in the principal’s interest to withhold any information in either case. This section investigates the potential merit of withholding information (about some dimension) by assuming that the states are imperfectly correlated, i.e., $0 < |\rho| < 1$. In this case, the agent’s marginal value of learning about one dimension is enhanced by excluding the other one because of substitutability.

First, I show that agent would acquire information for a smaller set of cost with some information being withheld than with full disclosure.

Proposition 6. The agent acquires information at a (weakly) larger range of cost when both dimensions are disclosed than when only one of the dimensions is disclosed. The inequality is strict if $0 < |\rho| < 1$.

When only the information about dimension $i$ is disclosed, agent would acquire information if and only if $c/2 \leq \sigma_i^2 \gamma_i^r + \sigma_j^2 \rho^2 \gamma_j^r$. But when both dimensions are disclosed, agent would acquire information for $c/2$ up to the prior variance $\Lambda_1$ of the first principal component. I show that the former threshold must be lower than $\Lambda_1$ if $0 < |\rho| < 1$. Therefore, when cost is sufficiently high, the principal should fully disclose both dimensions to enhance agent’s learning incentive.

Given that agent acquires information at a large cost only if all information is disclosed, it is natural to ask whether it could ever benefit the principal to withhold some information if agent still learns in this case. The next result shows that the principal should never withhold the dimension that is valued by her.

Proposition 7. For $\rho \neq \pm 1$, principal’s payoff from disclosing both dimensions is strictly higher than her payoff from disclosing only dimension two unless agent acquires no information.
in both regimes. For $\rho = \pm 1$, principal’s payoff from disclosing both dimensions is exactly equal to her payoff from disclosing only dimension two.

Recall that there are two underlying forces that influence principal’s payoff. On one hand, when being provided with more information, agent is able to formulate an attention strategy that is at least as well as, if not better than, having access to information about one dimension only. Because agent still values information about dimension one, which is given by the weight $\gamma_A^1 > 0$, the agent would attach a higher value to learning if he is able to freely construct his attention strategy with all information available. Hence agent is willing to devote more effort to learning overall when information is fully disclosed. On the other hand, the principal has a higher stake in the decision quality on dimension one relative to dimension two than the agent, which is given by $\gamma_P^1 / \gamma_P^2 \geq \gamma_A^1 / \gamma_A^2$. Since agent incurs loss from suboptimal actions on dimension one, however little it is, agent will acquire some information about this dimension if being provided. Although withholding information about dimension one would induce the agent to acquire more information about dimension two to compensate the loss, what agent learns instead could not substitute for the lost information about dimension one from the principal’s perspective. Therefore, principal should never withhold information about the dimension that she values more (than the agent).

It, however, could benefit the principal to exclude dimension two in certain conditions, because the effect of the two forces above now countervails each other. Restricting disclosure to dimension one only would induce agent to acquire less information overall but what is learned will be more valuable for the principal. Proposition 8 characterizes the condition under which exclusion could be beneficial to the principal (at some cost).

**Proposition 8.** The principal should withhold information about dimension two for some (medium) cost if and only if there exists $c > 0$ such that she gets a higher payoff from disclosing only dimension one than disclosing both dimensions when $c/2 = \Lambda_2$.

We only need to prove one direction of Proposition 8. Figure 3 illustrates that exclusion of information is most likely to be beneficial (for the principal) at the cutoff point $\Lambda_2$, principal’s payoff attains the highest possible value when $c = 0$, i.e., there is no cost, and is strictly higher than the payoff value when only either dimension is disclosed. But for $c/2$ less than $\Lambda_2$, principal’s payoff is decreasing at the highest rate when both dimensions are disclosed. It is possible that the blue line (“disclosing both dimensions”) crosses the red line (“disclosing dimension one only”) and principal gets a higher payoff by exclusion starting from there.

---

20 The marginal value of learning is high at minimal attention.

21 Agent will learn less to “compensate the loss” on dimension one than if principal herself acquires the information.

22 The horizontal axis of Figure 3 represents the cost parameter $c$, instead of $c/2$. So the two kink points of the blue line (“disclosing both dimensions”) are $2\Lambda_2$ and $2\Lambda_1$. 

23
show that the slope of the blue line will be flatter than the red line when $c/2$ is greater than $\Lambda_2$. Further, the two lines must cross again because the principal’s payoff with full disclosure is still positive when agent stops learning with exclusion. By Proposition 6 when cost is sufficiently large, agent only learns information if both dimensions are disclosed. The intuition about Proposition 8 is stated below Proposition 9 that contains analysis about agent’s learning activity in each dimension.

![Cost versus Principal Payoff](image)

**Figure 3:** $c$ versus $u^P$

**Corollary 2.** If $\gamma_1^P / \gamma_2^P = \gamma_1^A / \gamma_2^A$, then principal never benefits from excluding either dimension.

Corollary 2 confirms the idea that it is players’ misaligned interest in relative importance of multiple dimensions that creates potential benefit of exclusion. In contrast to Lester et al. (2012), the principal should never withhold any information if she attaches the same value as the agent to each dimension. Lester et al. (2012) assume that the jury (“agent”) has to process the whole piece of evidence to be able to use the information. Instead, the agent is free to choose any amount of information on each dimension in this model. If the jury does not incur a fixed cost in processing each piece of evidence but could choose the precision of interpretation (which implies different cost), I show that it is not in the principal’s interest to withhold any information (when they agree over the relative importance of each dimension).

Let us analyze each component of principal’s loss from the uncertainty in state realizations in more detail.

**Proposition 9.** Suppose that $\rho \neq \pm 1$. As cost increases, the posterior variance of dimension one is first lower and then higher when principal discloses only dimension one than when
principal discloses both dimensions. In contrast, the posterior variance of dimension two is always higher when principal only discloses dimension one than when principal discloses both dimensions.

![Change in Posterior Variances](image)

**Figure 4:** $c$ versus $\hat{\sigma}_1^2, \hat{\sigma}_2^2$

Proposition 9 holds no matter whether exclusion ever generates a higher payoff for the principal than full disclosure. The result is plotted in Figure 4. The benefit of exclusion lies in the fact that the principal and the agent trade off the reduction in uncertainty of the two dimensions in a different way.

When marginal cost $c$ is low enough, agent is willing to devote sufficient attention to learn both dimensions. Although agent learns (slightly) more about the first dimension if disclosure is restricted, principal would gain a higher payoff from agent’s much better knowledge about the realization of dimension two if there is no restriction. Hence principal should fully disclose both dimensions.

As marginal cost increases, agent learns less about both dimensions. Moreover, the further reduction in variance of dimension two $(1 - \rho^2)\sigma_2^2 - \left(\frac{c}{2\gamma_2}\right)(\frac{\gamma_2^A\sigma_2^2}{\gamma_1^A\sigma_1^2 + \gamma_2^A\rho^2\sigma_2^2})$, which is the gain from full disclosure, declines in cost $c$; while the additional uncertainty of dimension one $(\frac{c}{2\gamma_1})(\frac{\gamma_2^A\rho^2\sigma_2^2}{\gamma_1^A\sigma_1^2 + \gamma_2^A\rho^2\sigma_2^2})$, which is the loss from full disclosure, increases in $c$. Intuitively, as it becomes more costly to learn, the advantage of more learnable information from full disclosure shrinks. agent is not able to assimilate that much information even though being provided. So the gain from full disclosure would decline. The loss, however, grows for any $\rho \neq 0$, because the reduction in uncertainty about dimension one declines faster in cost when information about dimension two is available. In this case, agent learns less about two, which leads to less inference
about one. The additional layer of loss in information about dimension one is responsible for the increasing marginal loss from full disclosure. The net gain of full disclosure hence declines in cost initially. It will still be positive if they value information about the two dimensions in the same way. If principal values dimension one (relative to dimension two) more than agent, it is possible that the gain of a more informed decision on dimension two from full disclosure cannot compensate the loss of a worse decision on dimension one. It could then benefit principal to restrict disclosures in order to motivate agent to learn more about dimension one (at the expense of being less informed on dimension two though). The loss will be significant enough only if there is strong correlation between the two dimensions. When the correlation is not sufficiently strong, a less precise belief about dimension two by restriction has minimal impact on the marginal value of learning about dimension one and hence cannot enhance the incentive to learn more about dimension one.

When the cost parameter \( \frac{c}{2} \) goes beyond the threshold \( \Lambda_2 \)

\[
\begin{align*}
-(\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) + (\Lambda_1 - \frac{c}{2}) \left[ \frac{\gamma_1^P}{\gamma_1^A} (u_1^1)^2 + \frac{\gamma_2^P}{\gamma_2^A} (1 - (v_1^1)^2) \right]
\end{align*}
\]

if disclosing both dimensions; and

\[
\begin{align*}
-(\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) + (\gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2 - \frac{c}{2}) \left[ \frac{\gamma_1^A \sigma_1^2}{\gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2} + (\frac{\gamma_2^P}{\gamma_2^A})(1 - \frac{\gamma_1^A \sigma_1^2}{\gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2}) \right]
\end{align*}
\]

if disclosing dimension one only. Compared to lower cost, part of principal’s gain from full disclosure \( \Lambda_1 - \frac{c}{2} \)\left[ \frac{\gamma_1^P}{\gamma_1^A} (1 - (v_1^1)^2) + \frac{\gamma_2^P}{\gamma_2^A} (v_1^1)^2 \right] \) falls to zero (and cannot further decline).\footnote{Please see Lemma\ref{lemma:principal_payoff} for the result about principal’s payoff with full disclosure.} Recall that the learning outcome depends on the overall effort and the fraction assigned to each dimension. In the case of full disclosure, on one hand, agent is motivated to devote more effort, which can be seen from \( \Lambda_1 > \gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2 \). On the other hand, the fraction of dimension one is smaller and the fraction of dimension two is larger, i.e., \( (v_1^1)^2 < \frac{\gamma_1^A \sigma_1^2}{\gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2} \). It follows that agent learns strictly more about dimension two with full disclosure, but the advantage still continues to diminish. With regard to dimension one, the decreasing rate of information acquired is now slower if all information is fully disclosed. When \( \frac{c}{2} = \Lambda_2 \), agent learns more about dimension one if information is restricted, because the effect of higher fraction of attention for dimension one dominates. But as cost rises and less overall information is learned, the effect of the total effort induced would play a greater role. Proposition\ref{proposition:information_restriction} shows that finally the posterior variance of dimension one with restricted disclosure exceeds the variance with full disclosure. In particular, for \( \frac{c}{2} \geq \gamma_1^A \sigma_1^2 + \rho^2 \gamma_2^A \sigma_2^2 \), agent acquires information only if all is disclosed. Hence for cost that is high enough, it would not benefit principal to withhold either dimension of the information. In conclusion, principal withholds information if cost is in the medium range and there is sufficient correlation between the two dimensions.
We could think of the principal’s exclusion choice from another perspective. When principal withholds disclosure about a particular dimension, she is essentially raising agent’s attention cost to infinity on this dimension. So the exclusion strategy is the flip side of those in Section 5 in the sense that principal is influencing the agent’s cost now instead of agent’s benefit from learning (in reducing loss). Roughly, by increasing agent’s cost on his valued dimension, principal reduces the ratio of cost on her valued dimension relative to the agent’s. This superficial analogy implies that principal might gain from agent reallocating his attention as a cost-efficient decision, which is traded off against the decline in total effort elicited from the agent.

7 Volatility of the State

When the state is more volatile, principal’s information has more content and will be more useful for decision making of the agent. Failure to acquire the information, as a consequence of resistance to expend costly effort in learning, could generate a worse outcome in a more volatile environment. Sales managers need to learn information about markets of varying volatility. The juror evaluates evidence that differ in accuracy of the content. Students acquire knowledge that evolves and gets enriched over time. This section contains comparative statics predictions about principal’s payoff with respect to volatility of the state in each dimension, studying how the learning outcome will be affected by the amount of information from principal’s perspective.

If dimension one, that is more valued by the principal, becomes more volatile, I show that the principal will incur weakly more loss regardless of agent’s cost. Proposition 10 shows that principal’s payoff will not change for \( c/2 \leq \Lambda_2 \) or greater than \( \Lambda_1 \), because agent will be motivated to pay more attention to the increasingly volatile state if cost is small enough or resist learning anything if cost is big enough. For medium costs, however, the principal will become strictly worse off as \( \sigma_1^2 \) increases. It is true that agent is induced to acquire more information and allocate more attention to the dimension that contains more valuable information. The enhanced learning, nevertheless, still cannot offset the additional uncertainty, because principal cares about dimension one more than the agent and desires more learning (about this dimension) than what the agent did.

Proposition 10. Fix the variance of dimension two \( \sigma_2^2 \).

(i) When \( c/2 \leq \Lambda_2 \), principal’s payoff is constant as \( \sigma_1^2 \) increases.

(ii) When \( \Lambda_2 < c/2 \leq \Lambda_1 \), principal’s payoff is strictly decreasing in \( \sigma_1^2 \).

(iii) When \( c/2 > \Lambda_1 \), agent does not learn any information until \( \Lambda_1 \) reaches the threshold \( c/2 \) as \( \sigma_1^2 \) increases. The principal’s payoff will be first constant and then decreasing in \( \sigma_1^2 \) for \( \Lambda_1 \) greater than \( c/2 \).

Please see Section 8 for more discussion from this point of view.
If dimension two, that is more valued by the agent, becomes more volatile, principal’s payoff will still be constant for \( c/2 \) less than \( \Lambda_2 \) or greater than \( \Lambda_1 \), for the same reason as in Proposition 10. For \( c/2 \) between \( \Lambda_2 \) and \( \Lambda_1 \), a more volatile state in dimension two induces the agent to pay more attention but shift some fraction to dimension two. The overall effect would then depend on whether the rising effort or reallocation of attention dominates. Figure 5 shows a non-monotonic pattern between volatility of dimension two and principal’s payoff. The principal gains a higher utility for low values of \( \sigma_2^2 \), because the substitution effect is minimal when the (more valuable) information about the more uncertain state in dimension one draws the majority of agent’s attention.

**Proposition 11.** Fix the variance of dimension one \( \sigma_1^2 \).

(i) When \( c/2 \leq \Lambda_2 \), principal’s payoff is constant as \( \sigma_2^2 \) increases.

(ii) When \( \Lambda_2 < c/2 \leq \Lambda_1 \), the effect of change in \( \sigma_2^2 \) on principal’s payoff is indeterminate in general. The principal’s payoff can increase or decrease in \( \sigma_2^2 \).

(iii) When \( c/2 > \Lambda_1 \), agent does not learn any information until \( \Lambda_1 \) reaches the threshold \( c/2 \) as \( \sigma_2^2 \) increases.

![Figure 5: \( \sigma_2^2 \) versus \( u^P \)](image)

More accurately, how much (increasing) uncertainty in dimension two has been mitigated by the agent matters as well. Principal benefits from the enhanced precision about the decision on dimension one and suffers less than the agent from the additional uncertainty about dimension two that is not fully eliminated by agent.
8 Discussion

This paper explores the implication of rational inattention for disclosure about multiple issues. It shows that increasing agent’s incentive to learn his preferred information will be harmful from principal’s perspective only if the agent’s cost resides within a medium range. Instead, principal could benefit from excluding agent valued dimension for intermediate costs. The analysis is intended to be a block in building a theory about common practices in disclosure for judicial decision making, management of learning in organizations, and formulation of education policies. As an initial attempt, the model is selected to be the simplest one to illustrate intuitions, though several assumptions that potentially shape results are worth noting.

Either full disclosure or complete exclusion of some dimension is clearly a restriction, eliminating the possibility that the principal’s disclosure is imprecise but still informative about the dimension. It is straightforward to formulate a problem that the principal is able to choose an (intermediate) variance-covariance matrix and the agent cannot end up with a more precise belief than the principal’s disclosure. A major concern about this formulation is that it is not likely in the principal’s interest to disclose a Gaussian signal in general. But allowing for any arbitrary signal could render the solution intractable. My model, though restricted, delivers the (key) intuitions in the simplest way and does not touch on specific distributional assumptions about the principal’s signal.

A more pivotal modeling assumption is the agent being the only one who makes effort in communication. This is a conventional assumption in the persuasion and rational inattention literature, in which the agent’s cost is not internalized by the principal. On one hand, this model serves as a benchmark that studies strategic communication in the presence of rational inattention without imposing ad hoc assumptions about the principal’s signaling technology. On the other hand, it fails to capture an important feature of communication as a joint production process that requires efforts from both sides. A useful extension considers the situation in which principal’s and agent’s efforts are substitutable and principal could (invest to) deliver the materials in an easier way for the agent to understand. In other words, principal could (take actions to) manipulate agent’s attention cost, potentially inducing different cost for different dimensions. The discussion about excluding information in Section could be viewed as an extreme case in which the principal incurs no cost in communication and brings the agent’s cost to learn the undisclosed dimension to infinity. The result suggests that it could sometimes benefit the principal to (properly) increase agent’s cost (on some dimension), if it

---

26 The model described in the paper is a special case in which the principal influences the agent’s incentive through the information quality on each dimension.

27 The current strategies available to the principal lead to normally distributed signals.
is hard to directly mitigate cost.\textsuperscript{28} The effect on reallocation of the agent’s attention will be analogous to the case in which the principal reduces agent’s cost to process information about the disclosed dimension. It is not hard for us to model a technology available for the principal to share some agent’s cost given a plausible assumption on principal’s cost function.\textsuperscript{29} Presumably, it would benefit the principal to make some efforts to refine messages and improve disclosure quality, in order to relieve agent’s burden of interpretation and enhance understanding. Hence the exclusion strategy could be viewed as a second-best approach when it is not feasible for the principal to (directly) influence agent’s cognitive cost.

In addition, it could generate more insights by extending the model to an arbitrary $N$ dimensions of the state. In particular, if the principal incurs a cost as well in communication or could invest to reduce agent’s cost, it would be interesting to investigate how the principal should efficiently select a subset of issues to disclose. On the contrary, if the agent solely bears the cost of communication, the principal’s choice of dimensions might not be efficient, as exclusion of some dimension that is correlated to all other dimensions could force the agent to scrutinize more pieces of information.

\textsuperscript{28}The result in this extreme case implies that it would hold if principal \textit{optimally} increases the cost as well.
\textsuperscript{29}The exercise requires more than principal investing to generate a more precise signal that can be perceived by the agent. It is necessary to specify in the function of agent’s information cost how principal’s effort would translate to a lower cost for the agent. In the current model, an imprecise signal would only serve as a \textit{constraint} for agent’s potential posterior belief, without changing his cost structure.
References


Appendix

Lemma 1. Given a posterior belief about the state $x \sim N(\tilde{x}, \Sigma)$, the agent’s action $a_i$ is equal to the posterior mean $\tilde{x}_i$ for $i = 1, 2$.

Proof. The utility-maximizing actions are $\tilde{x}$, because certainty equivalence applies in a quadratic setup [Kőszegi and Matéjka 2020].

Proposition 1. The optimal information strategy is to acquire independent signals of $v^i \cdot w$ such that the posterior variance of $v^i \cdot w$ is $\min(c/2, \Lambda_i)$, where $v^i$ is the orthonormal basis of eigenvectors of $\Phi$ with the eigenvalue corresponding to $v^i$ denoted by $\Lambda_i$, and $w = (\sqrt{\gamma_1^A x_1}, \sqrt{\gamma_2^A x_2})$.

Proof. I first transform the agent’s problem in terms of the vector $x$ to a problem about the vector $w$ that captures agent’s utility weights.

(i) The objective functions are transferable:

$$\max_{\Sigma} -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{\lambda}{2} \log |\Sigma|$$

$$\iff \max_{\rho, \sigma_1, \sigma_2} -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{\lambda}{2} \log((\sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2)$$

$$\iff \max_{\rho, \sigma_1, \sigma_2} -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{\lambda}{2} \log((\sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2) + \log(\gamma_1^A \gamma_2^A))$$

$$\iff \max_{\rho, \sigma_1, \sigma_2} -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{\lambda}{2} \log((\gamma_1^A \gamma_2^A) \gamma_1^A \gamma_2^A)$$

$$\iff \max_{\rho, \sigma_1, \sigma_2} -\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + \frac{\lambda}{2} \log((\gamma_1^A \gamma_2^A) (\gamma_1^A \gamma_2^A) - (\sqrt{\gamma_1^A \gamma_2^A} \rho \sigma_1 \sigma_2)^2)$$

$$\iff \max_{z} -Tr(\Xi) + \frac{\lambda}{2} \log |\Xi|,$$

where $\Xi = \begin{pmatrix}
\gamma_1^A \sigma_1^2 & \sqrt{\gamma_1^A \gamma_2^A} \rho \sigma_1 \sigma_2 \\
\sqrt{\gamma_1^A \gamma_2^A} \rho \sigma_1 \sigma_2 & \gamma_2^A \sigma_2^2
\end{pmatrix}$.

(ii) The (no-forgetting) constraints are equivalent:

The “no-forgetting constraint” requires that $y^T(\Psi - \Sigma)y \geq 0$ for all non-zero column vector $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, i.e., $$(\sigma_1^2 - \sigma_1^2) y_1^2 + 2(\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2) y_1 y_2 + (\sigma_2^2 - \sigma_2^2) y_2^2 \geq 0.$$ It is equivalent to

$$(\sigma_1^2 - \sigma_1^2)(\sigma_2^2 - \sigma_2^2) - (\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2)^2 \geq 0 \quad \text{and} \quad \sigma_1^2 \leq \sigma_1^2.$$

I claim that $\Psi - \Sigma$ being positive semi-definite is equivalent to the following condition:

$$\Phi - \Xi = \begin{pmatrix}
\gamma_1^A (\sigma_1^2 - \sigma_1^2) & \sqrt{\gamma_1^A \gamma_2^A} (\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2) \\
\sqrt{\gamma_1^A \gamma_2^A} (\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2) & \gamma_2^A (\sigma_2^2 - \sigma_2^2)
\end{pmatrix}$$

is positive semi-definite,

where $\Phi = \begin{pmatrix}
\gamma_1^A \sigma_1^2 & \sqrt{\gamma_1^A \gamma_2^A} \rho \sigma_1 \sigma_2 \\
\sqrt{\gamma_1^A \gamma_2^A} \rho \sigma_1 \sigma_2 & \gamma_2^A \sigma_2^2
\end{pmatrix}$. Clearly, $\gamma_1^A (\sigma_1^2 - \sigma_1^2) y_1^2 + 2 \sqrt{\gamma_1^A \gamma_2^A} (\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2) y_1 y_2 + \gamma_2^A (\sigma_2^2 - \sigma_2^2) y_2^2 \geq 0$ for all non-zero vector $y$ requires that

$$(\gamma_1^A (\sigma_1^2 - \sigma_1^2)) \gamma_2^A (\sigma_2^2 - \sigma_2^2)) - (\sqrt{\gamma_1^A \gamma_2^A} (\rho \sigma_1 \sigma_2 - \rho \sigma_1 \sigma_2))^2 \geq 0 \quad \text{and} \quad \gamma_1^A (\sigma_1^2 - \sigma_1^2) \geq 0.$$
Because both $\gamma_1^A$ and $\gamma_2^A$ are nonnegative, the two conditions are equivalent.

So I transformed the problem into optimization over $\Xi$:

$$
\max_{\Xi} - Tr(\Xi) + \frac{\lambda}{2} \log |\Xi|
$$

subject to $\Phi - \Xi$ is positive semi-definite,

where $\Phi = \begin{pmatrix}
\gamma_1^A \sigma_1^2 & \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \\
\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} & \gamma_2^A \sigma_2^2
\end{pmatrix}$ and $\Xi = \begin{pmatrix}
\gamma_1^A \hat{\sigma}_1^2 & \sqrt{\gamma_1^A \gamma_2^A \rho \hat{\sigma}_1 \hat{\sigma}_2} \\
\sqrt{\gamma_1^A \gamma_2^A \rho \hat{\sigma}_1 \hat{\sigma}_2} & \gamma_2^A \hat{\sigma}_2^2
\end{pmatrix}$.

To explore what signals the agent would collect, I show that $\Xi$ should be diagonal in the basis of eigenvectors $v^i$. Decomposing the matrix $\Phi$ gives $\Phi = U\Lambda U'$, where $U$ is the unitary matrix with columns being the eigenvectors $v^1$ and $v^2$ such that $UU' = I$, and $\Lambda$ is a diagonal matrix with its elements $\Lambda_{ii}$ being the eigenvalues $\Lambda_i$ of $\Phi$. We can write the objective function as

$$
-Tr(\Xi) + \frac{\lambda}{2} \log |\Xi| = - Tr(UU' \Xi) + \frac{\lambda}{2} \log |UU' \Xi|
$$

$$
= - Tr(U' \Xi U) + \frac{\lambda}{2} \log |U||U'|||\Xi|
$$

$$
= - Tr(U' \Xi U) + \frac{\lambda}{2} \log |U'|||\Xi||U|
$$

$$
= - Tr(U' \Xi U) + \frac{\lambda}{2} \log |U'U||\Xi|
$$

$$
= - Tr(U' \Xi U) + \frac{\lambda}{2} \log |\Xi|
$$

$$
= - Tr(S) + \frac{\lambda}{2} \log |S|
$$

where $S \equiv U' \Xi U$ is the posterior variance-covariance matrix in the basis of eigenvectors of $\Phi$. Then $\Xi = (UU') \Xi (UU') = U(U' \Xi U) U' = USU'$. The “no-forgetting” constraint takes the form of $U'(\Phi - \Xi) U = U'(U\Lambda U')U - U'\Xi U = \Lambda - S$ being positive semi-definite. We then show that the optimal matrix $S$ is diagonal similar to K˝ oszegi and Matˇ ejka (2020).

Suppose that the optimal $S$ is not diagonal. Let $S^D$ be the diagonal matrix that has the same diagonal entries as $S$, i.e., $S^D_{ii} = S_{ii}$ for $i = 1, 2$ and $S^D_{12} = 0$.

(i) Because $\Lambda - S = \begin{pmatrix}
\Lambda_{11} - S_{11} & -S_{12} \\
-S_{12} & \Lambda_{22} - S_{22}
\end{pmatrix}$ is positive semi-definite, $S_{ii} \leq \Lambda_{ii}$ for $i = 1, 2$, which implies that $\Lambda - S^D = \begin{pmatrix}
\Lambda_{11} - S_{11} & 0 \\
0 & \Lambda_{22} - S_{22}
\end{pmatrix}$ is positive semi-definite as well.

(ii) Note that $-Tr(S^D) = -(S_{11} + S_{22}) = -Tr(S)$. So $S^D$ and $S$ generate the same loss.

(iii) But $|S| < S_{11}S_{22} = |S^D|$. So $\frac{\lambda}{2} \log |S| < \frac{\lambda}{2} \log |S^D|$, meaning that it would be more costly to acquire a signal having correlated dimensions given the same variances.

Hence the diagonal matrix $S^D$ gives agent a strictly higher payoff. The contradiction shows that the optimal posterior variance-covariance matrix has to be diagonal.
We can then find the posterior of each dimension separately, as the problem reduces to optimization over diagonal matrices:

\[
\max_{S_{11} \leq \Lambda_{11}, S_{22} \leq \Lambda_{22}} -S_{11} - S_{22} + \frac{\lambda}{2} \log S_{11} + \frac{\lambda}{2} \log S_{22}.
\]

Because the objective function is strictly concave over the entire domain of \(S_{ii} \ (i = 1, 2)\), I use the first-order approach to find the interior solution of the optimization problem. The first order condition with respect to \(S_{ii}\) gives \(-1 + \frac{\lambda}{2S_{ii}} = 0\). Since the function \(-s + \frac{\lambda}{2} \log s\) is strictly concave, the optimal matrix \(S\) with the constraint that posterior variances cannot exceed the prior variances is given by

\[
S = \begin{pmatrix}
\min(\frac{\lambda}{2}, \Lambda_{11}) & 0 \\
0 & \min(\frac{\lambda}{2}, \Lambda_{22})
\end{pmatrix}.
\]

Let \(M = (\gamma_2^2 \sigma_2^2 - \gamma_1^2 \sigma_1^2)/(2\sqrt{\gamma_1^2 \gamma_2^2 \rho \sigma_1 \sigma_2})\).

**Lemma 2.** The prior variances of the principal components are

\[
\Lambda_1 = \frac{1}{2} \left( \gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + \sqrt{(\gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2)^2 + (2\sqrt{\gamma_1^2 \gamma_2^2 \rho \sigma_1 \sigma_2})^2} \right);
\]

\[
\Lambda_2 = \frac{1}{2} \left( \gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 - \sqrt{(\gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2)^2 + (2\sqrt{\gamma_1^2 \gamma_2^2 \rho \sigma_1 \sigma_2})^2} \right).
\]

Furthermore, the eigenvectors satisfy \(v_1^2 = v_2^1\) and \(v_2^2 = -v_1^1\).

**Proof.** Recall that \(\Phi = \begin{pmatrix} \gamma_1 \sigma_1^2 & \sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2} \\ \sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2} & \gamma_2 \sigma_2^2 \end{pmatrix} = U\Lambda U^T\). The eigenvalues can be obtained immediately. Furthermore, simple calculations show that

\[
\begin{pmatrix} v_1^1 \\ v_2^1 \end{pmatrix} = \begin{pmatrix} 1 \\ \sqrt{1 + (\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2} - \sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2})^2} \end{pmatrix} \begin{pmatrix} \gamma_1^2 \sigma_1^2 \\ \gamma_2^2 \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \sigma_1^2 \\ \gamma_2 \sigma_2^2 \end{pmatrix} = \begin{pmatrix} v_1^2 \\ v_2^2 \end{pmatrix} = \begin{pmatrix} 1 \\ \sqrt{1 + (\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2} + \sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2})^2} \end{pmatrix} \begin{pmatrix} \gamma_1 \sigma_1^2 \\ \gamma_2 \sigma_2^2 \end{pmatrix} = \begin{pmatrix} v_1^2 \\ v_2^2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \sigma_1^2 \\ \gamma_2 \sigma_2^2 \end{pmatrix}.
\]

**Lemma 3.** The principal’s expected payoff with full disclosure is given by

\[
Eu^P = \begin{cases}
-c^2 \left( \frac{\gamma_1^2}{\gamma_1^2} + \frac{\gamma_2^2}{\gamma_2^2} \right) & \text{if } c/2 \leq \Lambda_2 \\
-(\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2) + (\Lambda_1 - c^2)(\frac{\gamma_1^2}{\gamma_1^2}(v_1^1)^2 + \frac{\gamma_2^2}{\gamma_2^2}(1-(v_1^1)^2)) & \text{if } \Lambda_2 < c/2 \leq \Lambda_1 \\
-(\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2) & \text{if } c/2 > \Lambda_1
\end{cases}
\]

\(30\) The expected return on the margin is the same for learning either dimension.
Lemma 5. This result immediately follows from Equation (2) and Lemma 2:

\[ u^P (\hat{\sigma}_1, \hat{\sigma}_2) = \gamma_1^P \sigma_1^2 - \gamma_2^P \sigma_2^2 \]

\[ = - \gamma_1^P (\Xi_{11}/\gamma_1^A) - \gamma_2^P (\Xi_{22}/\gamma_2^A) \]

\[ = - \frac{\gamma_1^P}{\gamma_1^A} (\min(\frac{c}{2}, \Lambda_1))(v_1)^2 + \min(\frac{c}{2}, \Lambda_2)(v_1)^2) + \frac{\gamma_2^P}{\gamma_2^A} (\min(\frac{c}{2}, \Lambda_1)(v_1)^2 + \min(\frac{c}{2}, \Lambda_2)(v_1)^2) \]

\[ = - \frac{\gamma_1^P}{\gamma_1^A} [\min(\frac{c}{2}, \Lambda_1)(v_1)^2 + \min(\frac{c}{2}, \Lambda_2)(1-(v_1)^2)] + \frac{\gamma_2^P}{\gamma_2^A} [\min(\frac{c}{2}, \Lambda_1)(1-(v_1)^2) + \min(\frac{c}{2}, \Lambda_2)(v_1)^2)] \]

\[ \begin{cases} 
- \frac{c}{2} \left( \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} \right) & \text{if } \frac{c}{2} \leq \Lambda_2 \\
- (\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) + (\Lambda_1 - \frac{c}{2}) [\frac{\gamma_1^P}{\gamma_1^A} (v_1)^2 + \frac{\gamma_2^P}{\gamma_2^A} (1-(v_1)^2)] & \text{if } \Lambda_2 < \frac{c}{2} \leq \Lambda_1 \\
- (\gamma_1^P \sigma_1^2 + \gamma_2^P \sigma_2^2) & \text{if } \frac{c}{2} > \Lambda_1 
\end{cases} \]

The following claims will be useful in proving Proposition 3.

Lemma 5. For \( 0 < |\rho| < 1, \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \geq \gamma_1^A \sigma_1^2 + (2\rho^2 - 1) \gamma_2^A \sigma_2^2 \).

Proof. The result is clear by observing that

\[ \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \geq \gamma_1^A \sigma_1^2 + (2\rho^2 - 1) \gamma_2^A \sigma_2^2 \]

\[ \Leftrightarrow (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho^2 \sigma_1^2 \sigma_2^2 \geq \gamma_1^A \sigma_1^2 + (2\rho^2 - 1) \gamma_2^A \sigma_2^2 \]

\[ \Leftrightarrow (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho^2 \sigma_1^2 \sigma_2^2 \geq (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho^2 \sigma_1^2 \sigma_2^2 - 4 \rho^2 (1 - \rho^2) (\gamma_2^A \sigma_2^2)^2 \]

\[ \Leftrightarrow 4 \rho^2 (1 - \rho^2) (\gamma_2^A \sigma_2^2)^2 > 0. \]

Lemma 6. For \( 0 < |\rho| < 1, - \frac{\partial \Lambda_2}{\partial \gamma_2^A} [\frac{\gamma_1^P}{\gamma_1^A} (1-(v_1)^2) + \frac{\gamma_2^P}{\gamma_2^A} (v_1)^2] + \frac{\gamma_2^P}{\gamma_2^A} [\frac{c}{2} (1-(v_1)^2) + \Lambda_2 (v_1)^2] > 0. \)

Proof. Note that

\[ - \frac{\partial \Lambda_2}{\partial \gamma_2^A} [\frac{\gamma_1^P}{\gamma_1^A} (1-(v_1)^2) + \frac{\gamma_2^P}{\gamma_2^A} (v_1)^2] + \frac{\gamma_2^P}{\gamma_2^A} [\frac{c}{2} (1-(v_1)^2) + \Lambda_2 (v_1)^2] \]

\[ \geq - \frac{\partial \Lambda_2}{\partial \gamma_2^A} \frac{\gamma_2^P}{\gamma_2^A} + \frac{\gamma_2^P}{\gamma_2^A} [\frac{c}{2} (1-(v_1)^2) + \Lambda_2 (v_1)^2] \]

\[ = \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{1}{\gamma_2^A} (1-(v_1)^2) + \Lambda_2 (v_1)^2 \right) - \frac{\partial \Lambda_2}{\partial \gamma_2^A} \]

\[ \geq \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{1}{\gamma_2^A} \Lambda_2 \right) = - \frac{\partial \Lambda_2}{\partial \gamma_2^A}. \]
where the first inequality follows from \( \frac{\sigma_i^2}{\gamma_1^2} \geq \frac{\gamma_i^2}{\gamma_2^2} \) and the second inequality follows from \( \Lambda_2 \leq c/2 \).

We show that \( \frac{\Lambda_2}{\gamma_2^2} > \frac{\partial \Lambda_2}{\partial \gamma_i^2} \), which will establish the proposition.

By definition and Eq (8),

\[
\frac{\Lambda_2}{\gamma_2^2} - \frac{\partial \Lambda_2}{\partial \gamma_i^2} = \left( \gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 \right) \sqrt{\left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2} - \left[ \left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2 \right]
\]

\[
2 \gamma_2^4 \sqrt{\left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2}
\]

\[
- \frac{\gamma_2^2 \sigma_2^2 \left( \gamma_1^4 \sigma_1^4 - \gamma_2^4 \sigma_2^4 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2 - \gamma_2^2 \sigma_2^2 + \gamma_1^4 \sigma_1^2 - 2 \gamma_1^4 \rho^2 \sigma_1^2}
\]

\[
2 \gamma_2^4 \sqrt{\left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2}
\]

The numerator is positive by Lemma 5. Hence \( \frac{\Lambda_2}{\gamma_2^2} > \frac{\partial \Lambda_2}{\partial \gamma_i^2} \), as claimed. \( \square \)

**Lemma 7.** For \( 0 < |\rho| < 1, \partial \Lambda_1/\partial \gamma_i^2 > 0 \) and \( \partial \Lambda_2/\partial \gamma_i^2 > 0 \) for \( i = 1, 2 \).

**Proof.** The derivative of \( \Lambda_1 \) with respect to \( \gamma_i^2 \) (\( i = 1, 2 \)) is

\[
\frac{\partial \Lambda_1}{\partial \gamma_i^2} = \frac{1}{2} \left[ \gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2 + \sqrt{\left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2} \right]
\]

\[
\sigma_2^2 \sqrt{\left( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 \right)^2 + 4 \gamma_j^4 \gamma_i^4 \rho^2 \sigma_j^2 \sigma_i^2} \left( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 \right) + 2 \gamma_j^4 \gamma_i^4 \rho^2 \sigma_j^2 \sigma_i^2)
\]

\[
\gamma_1^2 \sigma_1^2 \left[ \sqrt{\left( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 \right)^2 + 4 \gamma_j^4 \gamma_i^4 \rho^2 \sigma_j^2 \sigma_i^2} - \gamma_j^2 \sigma_j^2 + 2 \gamma_j^4 \rho^2 \sigma_j^2 \right]
\]

\[
2 \gamma_2^4 \sqrt{\left( \gamma_1^2 \sigma_1^2 - \gamma_2^2 \sigma_2^2 \right)^2 + 4 \gamma_1^4 \gamma_2^4 \rho^2 \sigma_1^2 \sigma_2^2}
\]

\[
\begin{cases}
\geq \sigma_i^2 \\
\frac{\sigma_i^2}{2} + \frac{\sigma_i^2}{2} + \sqrt{\left( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 \right)^2 + 4 \gamma_j^4 \gamma_i^4 \rho^2 \sigma_j^2 \sigma_i^2}
\end{cases}
\]

if \( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 + 2 \gamma_j^4 \rho^2 \sigma_j^2 \geq 0 \)

\[
\begin{cases}
\frac{\sigma_i^2}{2} + \frac{\sigma_i^2}{2} + \sqrt{\left( \gamma_j^2 \sigma_j^2 - \gamma_i^2 \sigma_i^2 \right)^2 + 4 \gamma_j^4 \gamma_i^4 \rho^2 \sigma_j^2 \sigma_i^2}
\end{cases}
\]

Otherwise.
If $\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2 + 2\gamma_j^2 \rho^2 \sigma_j^2 < 0$, the numerator is positive unless $|\rho| = 0$ or $1$. So $\Lambda_1$ is strictly increasing in $\gamma_i^2$ for $i = 1, 2$.

The derivative of $\Lambda_2$ with respect to $\gamma_i^2$ ($i = 1, 2$) is

$$
\frac{\partial \Lambda_2}{\partial \gamma_i^2} = \frac{1}{2} \frac{1}{\sqrt{\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2 + (2\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2})^2}}
$$

If $\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2 + 2\gamma_j^2 \rho^2 \sigma_j^2 > 0$, the numerator is positive unless $|\rho| = 0$ or $1$. So $\Lambda_2$ is strictly increasing in $\gamma_i^2$ for $i = 1, 2$.

Lemma 8. $\partial(v_1^2)/\partial \gamma_1^2 > 0$ and $\partial(v_1^2)/\partial \gamma_2^2 < 0$.

Proof. Recall that $M = (\gamma_1^2 \sigma_2^2 - \gamma_1^2 \sigma_1^2)/(2\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2})$.

$$
\frac{\partial(v_1^2)}{\partial \gamma_1^2} = \frac{\partial(v_1^2)}{\partial M} \frac{\partial M}{\partial \gamma_1^2}
$$

$$
= \left(\frac{\partial}{\partial M} \frac{1}{1 + (\sqrt{M^2 + 1} + M)^2}\right) \left(\frac{\partial}{\partial \gamma_1^2} \left(\frac{\sqrt{\gamma_2^2 \sigma_2^2} - \sqrt{\gamma_1^2 \sigma_1^2}}{2\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2}}\right)\right)
$$

$$
= - \frac{2(\sqrt{M^2 + 1} + M)(M/\sqrt{M^2 + 1} + M)}{[1 + (\sqrt{M^2 + 1} + M)^2]^2} \left[\frac{1}{2\rho} \frac{\sigma_1}{2\sqrt{\gamma_1 \gamma_2 \sigma_1}} + \frac{\sqrt{\gamma_1 \gamma_2 \sigma_1}}{2\sqrt{(\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2)}}\right] > 0.
$$

$$
\frac{\partial(v_1^2)}{\partial \gamma_2^2} = \frac{\partial(v_1^2)}{\partial M} \frac{\partial M}{\partial \gamma_2^2}
$$

$$
= \left(\frac{\partial}{\partial M} \frac{1}{1 + (\sqrt{M^2 + 1} + M)^2}\right) \left(\frac{\partial}{\partial \gamma_2^2} \left(\frac{\sqrt{\gamma_2^2 \sigma_2^2} - \sqrt{\gamma_1^2 \sigma_1^2}}{2\sqrt{\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2}}\right)\right)
$$

$$
= - \frac{2(\sqrt{M^2 + 1} + M)(M/\sqrt{M^2 + 1} + M)}{[1 + (\sqrt{M^2 + 1} + M)^2]^2} \left[\frac{1}{2\rho} \frac{\sigma_2}{2\sqrt{\gamma_1 \gamma_2 \sigma_2}} + \frac{\sqrt{\gamma_1 \gamma_2 \sigma_2}}{2\sqrt{(\gamma_1 \gamma_2 \rho \sigma_1 \sigma_2)}}\right] < 0.
$$

\[\square\]
Proposition 3. Fix agent’s value parameter on dimension two $\gamma_2^A$. When $c/2 \leq \Lambda_1$, the principal’s payoff is strictly increasing in $\gamma_1^A$. When $c/2 > \Lambda_1$, agent does not learn any information until $\Lambda_1$ reaches the threshold $c/2$ as $\gamma_1^A$ increases. The principal gains a higher payoff (than if no information is acquired) from here onwards.

Proof. (i) When $\Lambda_1 > \Lambda_2 > \gamma_2^A$, the principal’s payoff is equal to $-\gamma_1^A(\frac{\gamma^P}{\gamma_1^A} + \frac{\gamma^P}{\gamma_2^A})$. It is clear that if we marginally increase $\gamma_1^A$, principal’s payoff will increase. The uncertainty about dimension two will be the same, while the posterior variance of dimension one $\frac{\gamma^P}{\gamma_1^A}$ is reduced. If we increase $\gamma_1^A$, it will always be the case that $c/2 < \Lambda_2 < \Lambda_1$.

(ii) When $\Lambda_1 > \Lambda_2 > \Lambda_2$, agent only learns the direction associated with $\Lambda_1$. principal’s payoff is given by $-\gamma_2^A(\frac{\gamma^P}{\gamma_1^A} + \Lambda_2U_{11} + \frac{\gamma^P}{\gamma_2^A} + \Lambda_2U_{12})$. The sign of derivative with respect to $\gamma_2^A$ is generally indeterminate in this case. Nevertheless, we claim that principal’s payoff is locally strictly increasing in $\gamma_2^A$ if the relative value of learning about the two dimensions is the same for them.

(iii) When $\Lambda_2 < \Lambda_1 \leq \frac{\gamma^P}{\gamma_2^A}$, the principal’s payoff is equal to $-(\gamma_1^A\sigma_1^2 + \gamma_2^A\sigma_2^2)$. So principal’s payoff is constant in the marginal change of $\gamma_1^A$. \hfill \Box

Proposition 4. If $\frac{\gamma^P}{\gamma_2^A} = \frac{\gamma_1^A}{\gamma_2^A}$, then principal’s payoff is increasing in $\gamma_2^A$.

Proof. When $\Lambda_2 \leq \lambda/2 \leq \Lambda_1$, principal’s payoff is given by

$$ u_s = -\frac{\gamma_1^P}{\gamma_1^A} \left( \frac{\lambda}{2} U_{11} + \Lambda_2(1 - U_{11}^2) \right) + \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{\lambda}{2} (1 - U_{11}^2) + \Lambda_2 U_{11}^2 \right). $$

The derivative with respect to $\gamma_2^A$ is then given by

$$ \frac{\partial u_s}{\partial \gamma_2^A} = -\frac{\partial U_{11}}{\partial \gamma_2^A} \left( \frac{\gamma_1^P}{\gamma_1^A} - \frac{\gamma_2^P}{\gamma_2^A} \right) \left( \frac{\lambda}{2} - \Lambda_2 \right) $$

$$ -\frac{\partial \Lambda_2}{\partial \gamma_2^A} \left( \frac{\gamma_1^P}{\gamma_1^A}(1 - U_{11}^2) + \frac{\gamma_2^P}{\gamma_2^A} U_{11}^2 \right) + \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{\lambda}{2} (1 - U_{11}^2) + \Lambda_2 U_{11}^2 \right). $$

When $\frac{\gamma_1^P}{\gamma_1^A} = \frac{\gamma_1^A}{\gamma_1^A}$, $\frac{\gamma_1^P}{\gamma_1^A} = \frac{\gamma_1^A}{\gamma_1^A}$. So the first term of the right hand side of (8) is zero. We next show that the sum of the second and the third terms is positive.

It is clear that the third term is positive while the second term is negative because $\partial \Lambda_2/\partial \gamma_2^A > 0$ for $0 < |\rho| < 1$. When $\frac{\gamma_1^P}{\gamma_2^A} = \frac{\gamma_1^A}{\gamma_2^A}$,

$$ \frac{\partial u_s}{\partial \gamma_2^A} = -\frac{\partial U_{11}}{\partial \gamma_2^A} \left( \frac{\gamma_1^P}{\gamma_1^A} - \frac{\gamma_2^P}{\gamma_2^A} \right) \left( \frac{\lambda}{2} - \Lambda_2 \right) $$

$$ -\frac{\partial \Lambda_2}{\partial \gamma_2^A} \left( \frac{\gamma_1^P}{\gamma_1^A}(1 - U_{11}^2) + \frac{\gamma_2^P}{\gamma_2^A} U_{11}^2 \right) + \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{\lambda}{2} (1 - U_{11}^2) + \Lambda_2 U_{11}^2 \right) $$

$$ = -\frac{\partial \Lambda_2}{\partial \gamma_2^A} \left( \frac{\gamma_1^P}{\gamma_1^A}(1 - U_{11}^2) + \frac{\gamma_2^P}{\gamma_2^A} U_{11}^2 \right) + \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{\lambda}{2} (1 - U_{11}^2) + \Lambda_2 U_{11}^2 \right) > 0, $$

where the last inequality follows from Lemma 3. \hfill \Box
Lemma 4. The optimal information strategy if only dimension \(i\) is disclosed is to acquire the signal of \(i\) such that the posterior variance on dimension \(i\) is 
\[\sigma_i^2 \min\left(1, \frac{c}{2(\sigma_i^2 \gamma_i + \sigma_j^2 \rho^2 \gamma_j)}\right)\].

Proof. Let \(\Omega = \begin{pmatrix} \sigma_i & 0 \\ \sigma_j \rho & \sigma_j \sqrt{1 - \rho^2} \end{pmatrix}\). Then \(x = \Omega z\). Let \(\Upsilon\) be the posterior variance-covariance matrix of \(z\). Then \(\Psi = \Omega I \Omega^\prime\) and \(\Sigma = \Omega \Upsilon \Omega^\prime\). The cost incurred in learning is equivalent in terms of \(x\) and \(z\):

\[
\frac{\lambda}{2} \left( \log(|\Psi|) - \log(|\Sigma|) \right) = \frac{\lambda}{2} \left( \log(|\Omega I \Omega'|) - \log(|\Omega \Upsilon \Omega'|) \right)
\]

\[
= \frac{\lambda}{2} \left( \log(|\Omega||I||\Omega'|) - \log(|\Omega||\Upsilon||\Omega'|) \right)
\]

\[
= \frac{\lambda}{2} \left( \log(|\Omega| + \log|I| + \log|\Omega'| - (\log|\Omega| + \log|\Upsilon| + \log|\Omega'|) \right)
\]

\[
= -\frac{\lambda}{2} \log |\Upsilon|.
\]

Then the player’s loss from agent’s misperception is given by

\[
-\text{Tr} \left( \begin{pmatrix} \gamma_i & 0 \\ 0 & \gamma_j \end{pmatrix} \Sigma \right)
\]

\[
= -\text{Tr} \left( \begin{pmatrix} \gamma_i & 0 \\ 0 & \gamma_j \end{pmatrix} \Omega \Upsilon \Omega' \right)
\]

\[
= -\text{Tr} \left( \Omega' \begin{pmatrix} \gamma_i & 0 \\ 0 & \gamma_j \end{pmatrix} \Omega \Upsilon \right)
\]

\[
= -\text{Tr} \left( \begin{pmatrix} \sigma_i \gamma_i & \sigma_j \rho \gamma_j \\ \sigma_j \rho \gamma_j & \sigma_j \sqrt{1 - \rho^2} \gamma_j \end{pmatrix} \Omega \Upsilon \right)
\]

\[
= -\text{Tr} \left( \begin{pmatrix} \sigma_i^2 \gamma_i + \sigma_j^2 \rho^2 \gamma_j \\ \sigma_j^2 \rho \sqrt{1 - \rho^2} \gamma_j \end{pmatrix} \sigma_j^2 (1 - \rho^2) \gamma_j \right) \Upsilon \right).
\]

If principal only discloses the realization of dimension \(i\), agent’s signal cannot involve anything that is independent of this dimension. So the posterior variance of \(z_j\) is still 1. Moreover, \(z_i\) and \(z_j\) are still independent, as the signal about \(z_i\) reveals nothing about \(z_j\). Then \(\Upsilon\) is given by \(\hat{\sigma}_{z_i}^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\). The loss for players is

\[-(\sigma_i^2 \gamma_i + \sigma_j^2 \rho^2 \gamma_j) \hat{\sigma}_{z_i}^2 - \sigma_j^2 (1 - \rho^2) \gamma_j \].

Let the posterior variance of \(z_i\) be \(\hat{\sigma}_{z_i}^2\). Since agent’s cost is \(-\frac{\lambda}{2} \log |\Upsilon| = -\frac{\lambda}{2} \log \hat{\sigma}_{z_i}^2\), the agent’s problem is equivalent to

\[
\max_{\hat{\sigma}_{z_i}^2} - (\sigma_i^2 \gamma_i + \sigma_j^2 \rho^2 \gamma_j) \hat{\sigma}_{z_i}^2 - \sigma_j^2 (1 - \rho^2) \gamma_j + \frac{\lambda}{2} \log \hat{\sigma}_{z_i}^2
\]

subject to \(\hat{\sigma}_{z_i}^2 \leq 1\),
where $1$ is the prior variance of the random variable $z_i$.

The solution is given by the first order condition:

$$-(\gamma_i^2 \sigma_i^2 + \gamma_j^2 \rho^2 \gamma_j^2) + \left(\frac{\lambda}{2}\right)(\frac{1}{\sigma_{z_i}^2}) = 0.$$ 

So the optimal $\hat{\sigma}_{z_i}^2$ for the agent is $\min(1, \left(\frac{\lambda}{2(\gamma_i^2 \sigma_i^2 + \gamma_j^2 \rho^2 \gamma_j^2)}\right))$. \hfill \Box

**Proposition 6.** The agent acquires information at a (weakly) larger range of cost when both dimensions are disclosed than when only one of the dimensions is disclosed. The inequality is strict if $0 < |\rho| < 1$.

*Proof.* We compare the threshold cost in two cases, $2(\gamma_i^2 \sigma_i^2 + \gamma_j^2 \rho^2 \gamma_j^2)$ and $\gamma_i^2 \sigma_i^2 + \gamma_j^2 \rho^2 \gamma_j^2 + \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2}$. I claim that the threshold value when both dimensions are disclosed is strictly higher for $0 < |\rho| < 1$, which is shown below:

$$\left[\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2 + \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2} \right] > \left[2(\gamma_i^2 \sigma_i^2 + \gamma_j^2 \rho^2 \gamma_j^2)\right]$$

$$\Leftrightarrow \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2} > \gamma_i^2 \sigma_i^2 + (2\rho^2 - 1)\gamma_j^2 \sigma_j^2$$

$$\Leftrightarrow \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2} > \gamma_i^2 \sigma_i^2 + (2\rho^2 - 1)\gamma_j^2 \sigma_j^2$$

$$\Leftrightarrow (\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2 > (\gamma_i^2 \sigma_i^2)^2 + 2(2\rho^2 - 1)\gamma_i^2 \sigma_i^2 \gamma_j^2 \sigma_j^2 + (4\rho^2 - 4\rho^2 + 1)(\gamma_j^2 \sigma_j^2)^2$$

$$\Rightarrow 4\rho^2(1 - \rho^2)(\gamma_j^2 \sigma_j^2)^2 > 0.$$ 

The inequality above holds when $0 < |\rho| < 1$.

When $\rho = 0$, the threshold cost in the case of disclosure of both dimensions is $\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2 + \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2} = \gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2 + \gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2 \geq 2\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2 + (\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2) = 2\gamma_i^2 \sigma_i^2$, and the threshold cost when disclosing only dimension $i$ is $2\gamma_i^2 \sigma_i^2$.

When $\rho = \pm 1$, $\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2 + \sqrt{(\gamma_i^2 \sigma_i^2 - \gamma_j^2 \sigma_j^2)^2 + 4\gamma_i^2 \gamma_j^2 \rho^2 \sigma_i^2 \sigma_j^2} = 2(\gamma_i^2 \sigma_i^2 + \gamma_j^2 \sigma_j^2)$. Hence the threshold cost is the same in both cases. \hfill \Box

**Proposition 7.** For $\rho \neq \pm 1$, principal’s payoff from disclosing both dimensions is strictly higher than her payoff from disclosing only dimension two unless agent acquires no information in both regimes. For $\rho = \pm 1$, principal’s payoff from disclosing both dimensions is exactly equal to her payoff from disclosing only dimension two.

*Proof.* If $\rho = \pm 1$, $A_{22} = 0$ and $U_{11}^2 = \frac{\gamma_1^2 \sigma_1^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2}$. It is straightforward to check that principal’s payoff is identical in both regimes and equal to $-\frac{\lambda}{2}\left(\frac{\sigma_1^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2} + \frac{\sigma_2^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2}\right)$.

If $\rho \neq \pm 1$, we claim that principal’s payoff is always strictly higher when both dimensions are disclosed unless agent never acquires any information.

(i) If $\lambda \geq \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 2(\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2)^2}$, agent acquires no information in both regimes by Lemma ?? and principal’s payoff is the same.
(ii) If \(2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2) \leq \lambda < \gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}\), agent only acquires information if both dimensions are disclosed. So principal is strictly better off in this regime as the effect of learning \((\Lambda_{11} - \frac{\lambda}{2})(\frac{\gamma_1^P}{\gamma_1^4}U_{11}^2 + \frac{\gamma_2^P}{\gamma_2} (1 - U_{11}^2))\) adds to her payoff.

(iii) If \(\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 \leq \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2} \leq \lambda < 2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)\), agent learns in both regimes. In particular, only \(v^1\) is learned when both dimensions are disclosed. Note that principal’s payoff is integration of the increasing rate as cost declines. If only dimension two is disclosed, principal’s gain beyond her payoff with no information when unit cost is \(\lambda\) is given by

\[
\int_{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)}^{\lambda} \frac{\gamma_1^P \sigma_1^2 + \gamma_2^P \rho^2 \sigma_2^2}{2(\gamma_2^4\sigma_2^2 + \gamma_1^4\rho^2 \sigma_1^2)} dt
\]

\[
= \int_{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)}^{\lambda} \frac{1}{2} \frac{\gamma_1^P \sigma_1^2}{\gamma_2^4\sigma_2^2} \cdot \frac{\gamma_2^4\sigma_2^2}{\gamma_2^4\sigma_2^2 + \gamma_1^4\rho^2 \sigma_1^2} + \frac{\gamma_2^P \rho^2 \sigma_2^2}{\gamma_1^4\rho^2 \sigma_1^2} \cdot \frac{\gamma_1^4\rho^2 \sigma_1^2}{\gamma_2^4\sigma_2^2 + \gamma_1^4\rho^2 \sigma_1^2} dt
\]

\[
= \int_{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)}^{\lambda} \frac{\gamma_1^P}{\gamma_2^4} \cdot (1 - \frac{1}{\gamma_1^4\sigma_1^2 / \gamma_1^4\rho^2 \sigma_1^2} + 1) + \frac{\gamma_2^P}{\gamma_1^4} \cdot \frac{1}{\gamma_2^4\sigma_2^2 / \gamma_1^4\rho^2 \sigma_1^2} dt.
\]

If both dimensions are disclosed, principal’s gain beyond her payoff with no information when unit cost is \(\lambda\) is given by

\[
\int_{\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}}^{\lambda} \frac{\gamma_1^P}{\gamma_1^4} U_{11}^2 + \frac{\gamma_2^P}{\gamma_2^4} (1 - U_{11}^2) dt
\]

\[
= \int_{\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}}^{\lambda} \frac{\gamma_1^P}{\gamma_1^4} \cdot \frac{1}{1 + (\sqrt{\gamma_1^4\sigma_1^2} + M)^2} + \frac{\gamma_2^P}{\gamma_2^4} \cdot (1 - \frac{1}{1 + (\sqrt{\gamma_1^4\sigma_1^2} + M)^2}) dt,
\]

where \(M = (\gamma_2^4\sigma_2^2 - \gamma_1^4\sigma_1^2)/(2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})\). Then we unify the lower and upper limits, and the integrand differs only in the weights of \(\gamma_1^P/\gamma_1^4\) and \(\gamma_2^P/\gamma_2^4\). In particular, \((\sqrt{M^2 + 1} + M)^2 < \gamma_2^4\sigma_2^2 / \gamma_1^4\rho^2 \sigma_1^2\) for \(\rho^2 < 1\) by Lemma 11, i.e., the weight of the larger term \(\gamma_1^P/\gamma_1^4\) is higher in (3). It follows that

\[
\int_{\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}}^{\lambda} \frac{\gamma_1^P}{\gamma_1^4} U_{11}^2 + \frac{\gamma_2^P}{\gamma_2^4} (1 - U_{11}^2) dt
\]

\[
> \int_{\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}}^{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)} \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^4} \cdot \frac{1}{1 + (\sqrt{\gamma_1^4\sigma_1^2} + M)^2} + \frac{\gamma_2^P}{\gamma_2^4} \cdot (1 - \frac{1}{1 + (\sqrt{\gamma_1^4\sigma_1^2} + M)^2}) dt
\]

\[
> \int_{\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2 + \sqrt{(\gamma_1^4\sigma_1^2 - \gamma_2^4\sigma_2^2)^2 + (2\sqrt{\gamma_1^4\rho\sigma_1\sigma_2})^2}}^{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)} \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^4} \cdot \frac{1}{1 + (\sqrt{\gamma_1^4\sigma_1^2} / \gamma_1^4\rho^2 \sigma_1^2 + 1)^2} + \frac{\gamma_2^P}{\gamma_1^4} \cdot \frac{1}{\gamma_2^4\sigma_2^2 / \gamma_1^4\rho^2 \sigma_1^2 + 1} dt
\]

\[
= \int_{2(\rho^2\gamma_1^4\sigma_1^2 + \gamma_2^4\sigma_2^2)}^{\lambda} \frac{\gamma_1^P \sigma_1^2 + \gamma_2^P \rho^2 \sigma_2^2}{2(\gamma_2^4\sigma_2^2 + \gamma_1^4\rho^2 \sigma_1^2)} dt,
\]

44
because \( \gamma_1^P / \gamma_1^A \geq \gamma_2^P / \gamma_2^A \). Hence principal gains more from disclosing both dimensions when
\[
gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \leq \lambda < 2(\rho^2 \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2).
\]

(iv) When \( 0 \leq \lambda < \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \), the gain from disclosing both dimensions increases at two different intensities as cost declines, which is given by
\[
\int_\lambda^\gamma \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} dt
\]

\[
+ \int_\lambda^\gamma \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} (1 - U_{11}) dt
\]

\[
= \int_\lambda^\gamma \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} dt
\]

\[
+ \int_\lambda^\gamma \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} (1 - U_{11}) dt
\]

\[
= \int_\lambda^\gamma \frac{1}{2} \frac{\gamma_1^P}{\gamma_1^A} + \frac{\gamma_2^P}{\gamma_2^A} dt
\]

where the first inequality follows from Lemma 9 and the second inequality follows from Lemmas 9 and 11. If only dimension two is disclosed, principal’s gain beyond her payoff with no information is still given by
\[
\int_\lambda^\gamma \frac{\gamma_2^P}{\gamma_2^A} + \frac{\gamma_1^P}{\gamma_1^A} \rho^2 \sigma_1^2 dt
\]

\[
= \int_\lambda^\gamma \frac{\gamma_2^P}{\gamma_2^A} + \frac{\gamma_1^P}{\gamma_1^A} \rho^2 \sigma_1^2 dt
\]

It is then clear that principal gains strictly more from disclosing both dimensions. □

**Lemma 9.** As cost increases, principal’s expected payoff declines at a higher rate until \( \lambda \) is equal to \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \) when both dimensions are disclosed than when either one of the dimensions is disclosed.
Lemma 10. \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \leq 2(\sigma^2 \gamma^\rho + \sigma^2 \rho^2 \gamma^r) \) with strict inequality if \( \rho \neq 0 \).

Proof. Since \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \) is symmetric in \( \gamma_1^A \sigma_1^2 \) and \( \gamma_2^A \sigma_2^2 \).
\( \gamma_2^A \sigma_2^4 \), we write it as \( \gamma_1^i \sigma_i^2 + \gamma_j^j \sigma_j^2 - \sqrt{(\gamma_i^i \sigma_i^2 - \gamma_j^j \sigma_j^2)^2 + (2\sqrt{\gamma_i^i \gamma_j^j \rho \sigma_i \sigma_j})^2} \). Note that

\[
\gamma_1^i \sigma_i^2 + \gamma_j^j \sigma_j^2 - \sqrt{(\gamma_i^i \sigma_i^2 - \gamma_j^j \sigma_j^2)^2 + (2\sqrt{\gamma_i^i \gamma_j^j \rho \sigma_i \sigma_j})^2}
\leq \gamma_1^i \sigma_i^2 + \gamma_j^j \sigma_j^2 - |\gamma_i^i \sigma_i^2 - \gamma_j^j \sigma_j^2|
\leq \gamma_1^i \sigma_i^2 + \gamma_j^j \sigma_j^2 - (\gamma_i^i \sigma_i^2 - \gamma_j^j \sigma_j^2))
= 2\gamma_1^i \sigma_i^2
\leq 2(\sigma_i^2 \gamma_1^i + \sigma_j^2 \gamma_j^j).
\]

The first and the third inequalities above are strict when \( \rho \neq 0 \).

The following lemma is useful for proving Proposition 7.

**Lemma 11.** For \( \rho^2 < 1 \), \( (\sqrt{M^2 + 1} + M)^2 < \frac{\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2}{\gamma_1^A \rho \sigma_1 \sigma_2} \), where \( M = \frac{\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2}{2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2}} \).

**Proof.**

\[
(\sqrt{M^2 + 1} + M)^2 = \left( \frac{\gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2}{2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2}} + 1 \right) \left( \frac{\gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2}{2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2}} \right)
= \left( \frac{\gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2}{4\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} + \frac{\gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2}{2\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \right)^2
= \left( \frac{\gamma_2^A \sigma_2^2 + \gamma_1^A \sigma_1^2}{2\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} + \frac{\gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2}{2\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \right)^2
= \left( \frac{\gamma_2^A \sigma_2^2 + \gamma_1^A \sigma_1^2}{2\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2} \right)^2
= \frac{\gamma_2^A \sigma_2^2}{\gamma_1^A \rho \sigma_1 \sigma_2}.
\]

where the inequality follows from \( \rho^2 < 1 \).

**Lemma 12.** For \( \rho^2 < 1 \), \( (\sqrt{M^2 + 1} + M)^2 > \frac{\gamma_1^A \rho \sigma_1 \sigma_2}{\gamma_1^A \sigma_1^2} \), where \( M = \frac{\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2}{2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2}} \).

\[47\]
Proof. Note that
\[
(\sqrt{M^2+1}+M)^2 - \frac{\rho^2\sigma_1^2}{\gamma^2_1}\sigma_2^2
=\frac{((\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\rho^2\sigma^2_1\sigma^2_2+\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2}{2\sqrt{1+\gamma^2_1\rho\sigma_1\sigma_2}}
\]
\[
=\frac{((\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\rho^2\sigma^2_1\sigma^2_2+\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1+2\rho^2\sigma^2_2)}{2\sqrt{1+\gamma^2_1\rho\sigma_1\sigma_2}}
\]
\[
\cdot\frac{((\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\rho^2\sigma^2_1\sigma^2_2+\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1-2\rho^2\sigma^2_2)}{2\sqrt{1+\gamma^2_1\rho\sigma_1\sigma_2}}).
\]

The first term of the last equation must be positive, because \(\sqrt{(\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\gamma^2_1\gamma^2_2\rho^2\sigma^2_1\sigma^2_2} > \gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1\). The second term is positive as well, because for \(\rho \neq \pm 1\)
\[
\sqrt{(\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\gamma^2_1\gamma^2_2\rho^2\sigma^2_1\sigma^2_2} > 0
\]
\[
\iff (\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\gamma^2_1\gamma^2_2\rho^2\sigma^2_1\sigma^2_2 > (\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2-2\gamma^2_2\rho^2\sigma^2_1\sigma^2_2)\]
\[
\iff (\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2+4\gamma^2_1\gamma^2_2\rho^2\sigma^2_1\sigma^2_2 > (\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)^2-4\gamma^2_2\rho^2\sigma^2_2(\gamma^2_2\sigma^2_2-\gamma^2_1\sigma^2_1)+4(\gamma^2_2\rho^2\sigma^2_2)\]
\[
\iff 4(\gamma^2_2\rho\sigma_2^2)(\rho^2-1) < 0.
\]

\[\Box\]

Lemma 13. If \(\gamma^2_1\sigma^2_1+\gamma^2_2\sigma^2_2-\sqrt{(\gamma^2_1\sigma^2_1-\gamma^2_2\sigma^2_2)^2+2(\sqrt{1+\gamma^2_1\rho\sigma_1\sigma_2})} < \lambda < 2(\gamma^2_1\rho\sigma_1\sigma_2),\)
the decreasing rate of principal’s payoff in cost when disclosing both dimensions is strictly lower than the rate when disclosing only dimension one for \(\rho \neq \pm 1\).

Proof. The decreasing rate when disclosing both dimensions and when disclosing only dimension one are \(\frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} U^2_{11}+\frac{\gamma^2_2}{\gamma^2_2}(1-U^2_{11})\right)\) and \(\frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} \cdot \frac{\gamma^2_1}{\gamma^2_1+\gamma^2_2} \cdot \rho^2\sigma_1\sigma_2 + \frac{\gamma^2_2}{\gamma^2_2} \cdot \frac{\gamma^2_2\rho^2\sigma^2_2}{\gamma^2_1+\gamma^2_2\rho^2\sigma^2_2}\right)\), respectively. Note that
\[
\frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} U^2_{11}+\frac{\gamma^2_2}{\gamma^2_2}(1-U^2_{11})\right) = \frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} \cdot \frac{1}{1+(\sqrt{M^2+1}+M)^2} + \frac{\gamma^2_2}{\gamma^2_2} \cdot \frac{1}{1+(\sqrt{M^2+1}+M)^2}\right)
\]
and
\[
\frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} \cdot \frac{\gamma^2_1\sigma^2_1}{\gamma^2_1+\gamma^2_2\rho^2\sigma^2_2} + \frac{\gamma^2_2}{\gamma^2_2} \cdot \frac{\gamma^2_2\rho^2\sigma^2_2}{\gamma^2_1+\gamma^2_2\rho^2\sigma^2_2}\right) = \frac{1}{2}\left(\frac{\gamma^2_1}{\gamma^2_1} \cdot \frac{1}{1+(\gamma^2_2\rho^2\sigma^2_2/\gamma^2_1\sigma^2_1)} + \frac{\gamma^2_2}{\gamma^2_2} \cdot \frac{1}{1+(\gamma^2_2\rho^2\sigma^2_2/\gamma^2_1\sigma^2_1)}\right).
\]
So we compare \((\sqrt{M^2+1}+M)^2\) and \(\gamma^2_2\rho^2\sigma^2_2/\gamma^2_1\sigma^2_1\), where \(M = (\gamma^2_2\rho^2\sigma^2_2/\gamma^2_1\sigma^2_1)/(2\sqrt{1+\gamma^2_2\rho\sigma_1\sigma_2})\).

By Lemma 12 \((\sqrt{M^2+1}+M)^2 > \frac{\gamma^2_2\rho^2\sigma^2_2}{\gamma^2_1\sigma^2_1}\). Since \(\frac{\gamma^2_1}{\gamma^2_2} \geq \frac{\gamma^2_1}{\gamma^2_1}\), the decreasing rate in the case of disclosing both dimensions is lower.

\[\Box\]

If \(\frac{\gamma^2_1}{\gamma^2_2} = \frac{\gamma^2_1}{\gamma^2_1}\), the decreasing rate is the same.
So whether principal should ever restrict to dimension one depends on her payoff when \( \lambda/2 = \Lambda_{22} \), i.e., an intermediate value of \( \lambda \). Principal’s payoff if disclosing both dimensions is

\[
-\frac{\lambda}{2} (\frac{\gamma_1^P}{\gamma_1} + \frac{\gamma_2^P}{\gamma_2});
\]

principal’s payoff if disclosing only dimension one is

\[
-\frac{\lambda}{2} (\frac{\gamma_1^P}{\gamma_1} + \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1 A_2^2)}) + \frac{\gamma_2^P}{\gamma_2} (1 - \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1 A_2^2)}) - \sigma_2^2 (1 - \rho^2) \gamma_2^P.
\]

**Proposition 8.** The principal should withhold information about dimension two for some (medium) cost if and only if there exists \( c > 0 \) such that she gets a higher payoff from disclosing only dimension one than disclosing both dimensions when \( c/2 = \Lambda_2 \).

**Lemma 14.** For \( \rho \neq \pm 1 \), \( \Lambda_{22} < (1 - \rho^2) \gamma_2^A \sigma_2^2 \).

**Proof.** \( 2 \Lambda_{22} \) is equal to \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2) \) and \( 2(1 - \rho^2) \gamma_2^A \sigma_2^2 \) is equal to \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - (\gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2) \). The lemma follows by noting that

\[
\gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2)} < \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - (\gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2)
\]

\( \iff \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2)} > \gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2, \)

and the last inequality follows from Lemma 5. \( \square \)

**Corollary 2.** If \( \gamma_1^P / \gamma_2^P = \gamma_1^A / \gamma_2^A \), then principal never benefits from excluding either dimension.

**Proof.** When \( \rho = \pm 1 \), principal’s payoff is the same from disclosing both dimensions or disclosing only one dimension. When \( \rho \neq \pm 1 \), we prove that the difference in payoffs is positive when \( \lambda/2 = \Lambda_{22} \). It then follows that principal always gains a strictly higher payoff by revealing both dimensions.

By Lemma 14, \( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2)} < \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2 \). So the difference in payoffs is greater than

\[
-\frac{1}{2} (\gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - [\gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2])
\]

\[
\left(\frac{\gamma_1^P}{\gamma_1} (1 - \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1 A_2^2)}) + \frac{\gamma_2^P}{\gamma_2} \right) + \frac{\gamma_2^P}{\gamma_2} (1 - \rho^2)
\]

\[
= -\frac{1}{2} (\gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - [\gamma_1^A \sigma_1^2 + (2 \rho^2 - 1) \gamma_2^A \sigma_2^2]) + \frac{\gamma_2^P}{\gamma_2} (1 - \rho^2)
\]

\[
= - (1 - \rho^2) \gamma_2^A \sigma_2^2 (\frac{\gamma_2^P}{\gamma_2}) + \frac{\gamma_2^P}{\gamma_2} (1 - \rho^2)
\]

\[
= - (1 - \rho^2) \gamma_2^A \sigma_2^2 + \gamma_2^P (1 - \rho^2) = 0,
\]

where the first equality follows from \( \gamma_1^P / \gamma_2^P = \gamma_1^A / \gamma_2^A \), which implies that \( \gamma_1^P / \gamma_1^A = \gamma_2^P / \gamma_2^A \). \( \square \)
Proposition 9. Suppose that \( \rho \neq \pm 1 \). As cost increases, the posterior variance of dimension one is first lower and then higher when principal discloses only dimension one than when principal discloses both dimensions. In contrast, the posterior variance of dimension two is always higher when principal only discloses dimension one than when principal discloses both dimensions.

Proof. When \( \lambda/2 \leq \Lambda_{22} \), the posterior variances of the two dimensions if principal disclose both are given by \( \frac{\lambda}{2\gamma_2} \) and \( \frac{\lambda}{2\gamma_1} \), respectively. The posterior variances are instead \( \left( \frac{\lambda}{2\gamma_1} \right) \left( \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right) \) and \( \left( \frac{\lambda}{2\gamma_2} \right) \left( 1 - \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right) + (1 - \rho^2)\sigma_2^2 \), respectively, if principal only discloses dimension one. It is clear that the variance of the first dimension would be lower if principal only discloses dimension one. For the second dimension, note that

\[
\left( \frac{\lambda}{2\gamma_2} \right) \left( 1 - \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right) + (1 - \rho^2)\sigma_2^2
\]

\[
= \frac{\lambda}{2\gamma_2} - (1 - \rho^2)\sigma_2^2 - \left( \frac{\lambda}{2\gamma_2} \right) \left( \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right)
\]

\[
\geq \frac{\lambda}{2\gamma_2} - (1 - \rho^2)\sigma_2^2 - \left( \frac{\lambda}{2\gamma_2} \right) \left( \frac{\Lambda_{22}}{\gamma_2^A} \right) \left( \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right)
\]

\[
> \frac{\lambda}{2\gamma_2} - (1 - \rho^2)\sigma_2^2 - \left( \frac{1}{\gamma_2^A} \right) \left( \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right)
\]

\[
= \frac{\lambda}{2\gamma_2} - (1 - \rho^2)\sigma_2^2 \left[ 1 - \frac{1}{1 + (\gamma_2^A \rho^2 \sigma_2^2 / \gamma_1^A \sigma_1^2)} \right]
\]

\[
> \frac{\lambda}{2\gamma_2}.
\]

So the posterior variance of the second dimension is higher if principal discloses only dimension one.

When \( \gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2 \leq \lambda/2 \leq \Lambda_{11} \), agent acquires information only if principal discloses both dimensions. Since \( 0 < U_{11}^2 < 1 \), the variance of both dimensions will be lower when principal discloses all information.

When \( \Lambda_{22} < \lambda/2 < \gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2 \), we first consider the case in which principal discloses both dimensions. The reduction in variance of dimension one is given by

\[
\int_{\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}}^{\lambda} \frac{1}{2\gamma_1^A} U_{11}^2 dt - \frac{1}{2\gamma_1^A} U_{11}^2 dt
\]

\[
= \int_{\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}}^{\lambda} \frac{1}{2\gamma_1^A} U_{11}^2 dt
\]

\[
= \int_{\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2}^{2(\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2)} \frac{1}{2\gamma_1^A} U_{11}^2 dt + \int_{2(\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2)}^{\gamma_1^A \sigma_1^2 + \gamma_2^A \rho^2 \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}} \frac{1}{2\gamma_1^A} U_{11}^2 dt;
\]

50
while the reduction in variance of dimension two is given by

\[
\begin{align*}
\int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 & \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2} \\
& = \int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2} \\
& > \int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2} \\
& > \int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2},
\end{align*}
\]

where the first inequality follows from Lemma 12 and the second inequality follows from Lemma 12.

If principal discloses only dimension one, the reduction in variance of dimension one is given by

\[
\int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2};
\]

while the reduction in variance of dimension two is given by

\[
\int_{\lambda}^2 (\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2) - 1 \overbrace{2 \gamma_2^4 (1 - U_{11}^2) dt}^{(1) - U_{11}^2}.
\]

So the variance of dimension two must be higher if principal discloses only dimension one. The variance of dimension one will be lower when disclosing only dimension one if \(\lambda\) is relatively small but higher if \(\lambda\) is relatively large. To see this, when \(\gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2 \leq \lambda/2 \leq \Lambda_{11}\), the variance is higher, because agent does not acquire any information if disclosing only dimension one. As cost decreases, agent starts to acquire information in both cases. Although the increasing rate \((\frac{1}{2 \gamma_1^4})(\frac{1}{1 + \gamma_2^4 \rho^2 \sigma_2^2 / \gamma_1^4 \sigma_1^2})\) is higher than the rate \((\frac{1}{2 \gamma_1^4})U_{11}^2\) by Lemma 12, the variance has already been reduced by \(\frac{1}{1 + \gamma_2^4 \rho^2 \sigma_2^2 / \gamma_1^4 \sigma_1^2}\) when \(\lambda/2 = \gamma_1^4 \sigma_1^2 + \gamma_2^4 \rho^2 \sigma_2^2\) if principal discloses both dimensions. Hence initially, the variance is higher if restricting to dimension one. The variance changes continuously and finally there is less remaining uncertainty with the restriction because when \(\lambda/2 = \Lambda_{22}\), the posterior variance of dimension one \((\frac{\Delta_{22}}{\gamma_1^4})(\frac{1}{1 + (\gamma_2^4 \rho^2 \sigma_2^2 / \gamma_1^4 \sigma_1^2)})\) is less than \(\frac{\Delta_{22}}{\gamma_1^4}\). This completes the proof. \(\square\)

**Proposition 10.** Fix the variance of dimension two \(\sigma_2^2\).

(i) When \(c/2 \leq \Lambda_2\), principal’s payoff is constant as \(\sigma_1^2\) increases.

(ii) When \(\Lambda_2 < c/2 \leq \Lambda_1\), principal’s payoff is strictly decreasing in \(\sigma_1^2\).

(iii) When \(c/2 > \Lambda_1\), agent does not learn any information until \(\Lambda_1\) reaches the threshold \(c/2\) as \(\sigma_1^2\) increases. The principal’s payoff will be first constant and then decreasing in \(\sigma_1^2\) for \(\Lambda_1\) greater than \(c/2\).
Proof. (i) When $\lambda/2 < \Lambda_{22} < \Lambda_{11}$, principal’s payoff is given by $-(\lambda/2)[(\gamma_1^P/\gamma_1^A) + (\gamma_2^P/\gamma_2^A)]$, which is invariant with respect to $\sigma_1^2$. Furthermore,

$$\frac{\partial \Lambda_{11}}{\partial \sigma_1^2} = \frac{1}{2} \left[ \left( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right) \right]$$

$$= \frac{1}{2} \left[ \frac{\gamma_1^A}{\gamma_2^A} \left( \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right) \right]$$

$$= \left( \frac{\gamma_1^A}{\gamma_2^A} \right) \left( \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right)$$

$$= \left( \frac{\gamma_1^A}{\gamma_2^A} \right) \left( \frac{\sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right)$$

It is clear that for $0 < |\rho| < 1$,

$$= (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho^2 \sigma_1^2 \sigma_2^2 - 4 (\gamma_2^A \sigma_2^2)^2 \rho^2 (1 - \rho^2)$$

$$= (\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + 4 \gamma_1^A \gamma_2^A \rho^2 \sigma_1^2 \sigma_2^2$$

Then $\partial \Lambda_{11}/\partial \sigma_1^2 > 0$ for $0 < |\rho| < 1$. Similarly,

$$\frac{\partial \Lambda_{22}}{\partial \sigma_1^2} = \frac{1}{2} \left[ \left( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 - \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right) \right]$$

$$= \left( \frac{\gamma_1^A}{\gamma_2^A} \right) \left( \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right)$$

$$= \left( \frac{\gamma_1^A}{\gamma_2^A} \right) \left( \frac{\sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2 \sqrt{\gamma_1^A \gamma_2 \rho \sigma_1 \sigma_2})^2} \right)$$

It follows from the same analysis as above that $\partial \Lambda_{22}/\partial \sigma_1^2 > 0$ for $0 < |\rho| < 1$. Therefore principal’s payoff will be constant however $\sigma_1^2$ increases.

(ii) When $\Lambda_{22} \leq \lambda/2 \leq \Lambda_{11}$, principal’s payoff does depend on the value of $\sigma_1^2$. The derivative of principal’s payoff with respect to $\sigma_1^2$ is

$$\frac{\partial}{\partial \sigma_1^2} \left[ \gamma_1^P \left( \frac{\lambda}{2} U_{11} + \Lambda_{22}(1 - U_{11}) \right) + \frac{\gamma_2^P}{\gamma_2^A} \left( \frac{\lambda}{2} (1 - U_{11}) + \Lambda_{22} U_{11} \right) \right]$$

$$= - \frac{\partial \Lambda_{22}}{\partial \sigma_1^2} \left[ \gamma_1^P \left( 1 - U_{11}^2 \right) + \frac{\gamma_2^P}{\gamma_2^A} U_{11}^2 + \frac{\partial U_{11}}{\partial \sigma_1^2} \left( \frac{\lambda}{2} - \Lambda_{22} \right) \left( \frac{\gamma_1^P}{\gamma_1^A} - \frac{\gamma_2^P}{\gamma_2^A} \right) \right]$$

52
Since $\frac{\partial \Lambda_{22}}{\partial \sigma_1^2} > 0$, the first term inside the bracket $\frac{\partial \Lambda_{22}}{\partial \sigma_1^2}(\gamma_1^p(1 - U_{11}^2) + \gamma_2^p U_{11}^2)$ is positive. We next show that $\frac{\partial \sigma_1^2}{\partial \sigma_1^2} > 0$:

$$\frac{\partial U_{11}^2}{\partial \sigma_1^2} = \frac{\partial U_{11}^2}{\partial M} \cdot \frac{\partial M}{\partial \sigma_1^2}$$

$$= \frac{\partial U_{11}^2}{\partial M} \left(\frac{\partial M}{\partial \sigma_1} \cdot \frac{1}{\partial \sigma_1^2}\right)$$

$$= \left(\frac{1}{2\sigma_1}\right) \left(\frac{\partial}{\partial \sigma_1^2} \left[\left(\frac{M}{\sqrt{M^2 + 1}} + 1\right)\left(\frac{\gamma_1^p}{\gamma_1^2} - \frac{\gamma_2^p}{\gamma_2^2}\right)\right]\right)$$

$$= - \left(\frac{1}{2\sigma_1}\right) \left(\frac{2(M^2 + 1 + M)(M/\sqrt{M^2 + 1} + 1)}{[1 + (\sqrt{M^2 + 1} + 1)]^2}\right) \left(\frac{1}{2\rho}\right) \left(\frac{-\sqrt{\gamma_1^2 \gamma_2^2 - \gamma_1^2 \gamma_2^2}}{\gamma_1^2 \gamma_2^2}\right) > 0.$$ 

Further, $\frac{\partial}{\partial \sigma_1} > \Lambda_{22}$ in this case and $\frac{\gamma_1^p}{\gamma_1^2} - \frac{\gamma_2^p}{\gamma_2^2} \geq 0$ by Assumption 1. Hence the second term inside the bracket is nonnegative as well. It follows that principal’s payoff is strictly decreasing in $\sigma_1^2$.

(iii) When initially $\lambda_0 > \Lambda_{11} > \Lambda_{22}$, agent does not learn anything. So principal’s payoff will be constant at the level without any information until $\Lambda_{11}$ hits $\lambda_0$. Because $\Lambda_{11}$ and $\Lambda_{22}$ are strictly increasing in $\sigma_1^2$, a higher $\sigma_1^2$ will motivate agent to learn once $\Lambda_{11}$ exceeds $\lambda_0$. But because principal cares more about the first dimension than agent and desires more learning (about this dimension) than what agent did, the enhanced learning cannot countervail the increase in $\sigma_1^2$ and principal is strictly worse off. Finally, $\Lambda_{22}$ hits $\lambda_0$ and principal’s payoff is constant at $-\lambda_0(\gamma_1^p/\gamma_1^2 + \gamma_2^p/\gamma_2^2)$. \qed

**Proposition 11.** Fix the variance of dimension one $\sigma_1^2$.

(i) When $c/2 \leq \Lambda_2$, principal’s payoff is constant as $\sigma_2^2$ increases.

(ii) When $\Lambda_2 < c/2 \leq \Lambda_1$, the effect of change in $\sigma_2^2$ on principal’s payoff is indeterminate in general. The principal’s payoff can increase or decrease in $\sigma_2^2$.

(iii) When $c/2 > \Lambda_1$, agent does not learn any information until $\Lambda_1$ reaches the threshold $c/2$ as $\sigma_2^2$ increases.

**Proof.** When there is more information content in the second dimension, principal’s payoff could increase when the increase in effort outweighs the substitution effect.

(i) When cost is low enough (relative to $\sigma_2^2$), i.e., $\lambda_0 > \Lambda_{22} < \Lambda_{11}$, how much and what information agent acquires does not depend on the prior and hence principal’s payoff is constant at $-(\lambda_0/2)[(\gamma_1^p/\gamma_1^2) + (\gamma_2^p/\gamma_2^2)]$. Similarly, $\partial \Lambda_{11}/\partial \sigma_2^2$ and $\partial \Lambda_{22}/\partial \sigma_2^2$ are both positive from the analysis above by simply replacing the role of $\sigma_1^2$ with $\sigma_2^2$. So principal’s payoff will be constant if $\sigma_2^2$ increases.

(ii) When $\Lambda_{22} < \lambda_0 < \Lambda_{11}$, only the principal component associated with $\Lambda_{11}$ will be
learned. The derivative of principal’s payoff with respect to $\sigma^2$ is given by

$$\frac{\partial}{\partial \sigma^2} \left[ -(\gamma_P^1 \sigma_1^2 + \gamma_2^P \sigma_2^2) + (\Lambda_{11} - \frac{\lambda}{2})(\frac{\gamma_1^P}{\gamma_1^A}U_{11}^2 + \frac{\gamma_2^P}{\gamma_2^A}(1 - U_{11}^2)) \right]$$

$$= -\gamma_2^P + \frac{\partial \Lambda_{11}}{\partial \sigma^2} \left( \frac{\gamma_1^P}{\gamma_1^A}U_{11}^2 + \frac{\gamma_2^P}{\gamma_2^A}(1 - U_{11}^2) \right) + \frac{\partial U_{11}^2}{\partial \sigma^2} (\Lambda_{11} - \frac{\lambda}{2})(\frac{\gamma_1^P}{\gamma_1^A} - \frac{\gamma_2^P}{\gamma_2^A})$$

Gain from learning more

Loss from substituting effort

Similarly, we can show that $\frac{\partial \Lambda_{11}}{\partial \sigma^2} > 0$ and $\frac{\partial U_{11}^2}{\partial \sigma^2} < 0$.

$$\frac{\partial \Lambda_{11}}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \left[ \frac{1}{2} \left( \gamma_1^A \sigma_1^2 + \gamma_2^A \sigma_2^2 + \sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} \right) \right]$$

$$= \frac{1}{2} \frac{\gamma_1^A}{\sqrt{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}} \left( \gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 \right)$$

$$= \frac{\gamma_1^A}{2} (\frac{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2})$$

$$\geq \frac{\gamma_1^A}{2} \frac{\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2}{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2}$$

if $\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + 2\gamma_1^A \rho \sigma_1 \sigma_2 \geq 0$

$$= \frac{\gamma_1^A}{2} \frac{\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2}{(\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2)^2 + (2\sqrt{\gamma_1^A \gamma_2^A \rho \sigma_1 \sigma_2})^2} > 0$$

if $\gamma_1^A \sigma_1^2 - \gamma_2^A \sigma_2^2 + 2\gamma_1^A \rho \sigma_1 \sigma_2 < 0$.

$$\frac{\partial U_{11}^2}{\partial \sigma^2} = \frac{\partial U_{11}^2}{\partial M} \cdot \frac{\partial M}{\partial \sigma^2}$$

$$= \frac{\partial U_{11}^2}{\partial \sigma^2} \left( \frac{\partial M}{\partial \sigma^2} \right)$$

$$= \frac{1}{2\sigma^2} \left( \frac{\partial}{\partial M} \left( \frac{1}{\sqrt{M^2 + 1 + M}} \right) \right) \left( \frac{\partial}{\partial \sigma^2} \left( \frac{\sqrt{\gamma_2^A \sigma_2}}{2 \sqrt{\gamma_1^A \rho \sigma_2}} - \frac{\sqrt{\gamma_1^A \sigma_1}}{2 \gamma_2^A \rho \sigma_2} \right) \right)$$

$$= -\frac{1}{2\sigma^2} \left( \frac{2(M^2 + 1 + M)(M/\sqrt{M^2 + 1 + M})}{[1 + (\sqrt{M^2 + 1 + M})^2]} \right) \left( \frac{1}{2\rho} \left( \frac{\sqrt{\gamma_2^A \sigma_2}}{\sqrt{\gamma_1^A \sigma_1}} + \frac{\sqrt{\gamma_1^A \sigma_1}}{\sqrt{\gamma_2^A \sigma_2}} \right) \right) < 0.$$
parameter $\lambda$. Again, the gain for principal because of more effort spent by agent would dominate when cost is sufficiently large and there is initially too limited information having been learned that is given by $\Lambda_{11} - \lambda/2$. Moreover, the gain from learning more does not depend on cost (constant in cost).

(iii) When $\lambda/2 > \Lambda_{11} > \Lambda_{22}$, principal’s payoff will be constant at $-\left(\gamma_1 P_1 \sigma_1^2 + \gamma_2 P_2 \sigma_2^2\right)$. As $\sigma_2^2$ increases, agent will start learning when $\Lambda_{11}$ is equal to $\lambda/2$. \hfill \Box