A financial crisis is often followed by a technological or a financial innovation. Recent examples include the dot-com bubble and the global financial crisis of 2008 which involved innovative financial products such as collateralized debt obligations (CDOs). One of the explanations for why a financial crisis occurs is that economic agents are irrational and ignore the tail risks in their investment choices because of various biases such as the availability heuristic or ``blindness to outliers".

However, the financial industry is known to hire talented individuals. Therefore, it is surprising that these talented individuals make systematic mistakes, particularly in an environment where there are strong incentives to make large profits if their portfolios perform well. In this paper, we answer this puzzle and provide an explanation for why rational agents may choose to learn less about the left tail than the right tail of the distribution of investments they make. As a consequence, they end up investing in even those assets that have a fat left tail and therefore a high likelihood of failure ex post.

We build a principal-agent model with flexible information acquisition. The principal (``she") wants to make an investment and has two choices, an asset A with unknown returns about which costly information can be acquired to learn about its distribution or another asset B about which there is nothing to learn. However, the principal does not have the skill to acquire the information about asset A, so she hires an agent (“he”) to do so. Since an agent can choose both the intensity and nature of information acquisition, we use the approach of flexible information acquisition to model our information structure. Flexible information acquisition implies that the agent can choose any possible information structure which allows us to model state-contingent attention allocation. Since information acquisition is costly, the contract offered to the agent determines how much effort he exerts in learning about different states. Thus our approach allows us to endogenously model differential attention allocation to different states by the agent and the optimal contract that the principal offers. We model information cost as the reduction in Shannon's entropy of a random variable. The key friction in our model is that the nature and intensity of information acquired by the agent is unobservable and that the agent has limited liability.

In the benchmark (first-best) case, the principal herself exerts effort to acquire information about asset A. In the benchmark information structure, the principal acquires information about all values in the support of the distribution of cash flows from asset A. This is because she fully internalizes the profit (or loss) from each possible outcome of cash flow generated by asset A.

However, if the agent acquires the information (second-best), then less information is acquired about the left side of the distribution of asset A compared to the first-best. In the optimal contract, if the agent chooses asset A, then for low cash flows the agent is paid zero. If the cash flow is above a certain cut-off, then the agent is paid an increasing concave wage. To incentivize the agent to acquire information about asset A and then choose the right project conditional on that information, the agent should be punished adequately if he chooses asset A and the return is very low. This punishment should be large enough relative to the expected payoff from choosing asset B. However, because of limited liability, the best the principal can do is give the agent zero if low outcomes occur, as opposed to giving him a negative wage. Given the flat nature of the contract at low returns from the asset A, the agent has no incentive to exert effort to acquire information to distinguish between these low outcomes. Hence he does not acquire any information about the left tail of asset A.

Hence the information intensity is zero in the region where the wage is zero. The consequence of this is that the posterior distributions after observing the signals are very close to the prior on the left side of the distribution, but much apart from the prior on the right side of the distribution. In other words, the agent learns very little about the left side of the distribution.