Review Article

Self-Confirming Biased Beliefs in Organizational “Learning by Doing”

Sanghyun Park and **Phanish Puranam**

Department of Strategy, INSEAD, Ayer Rajah Avenue, Singapore

Correspondence should be addressed to Phanish Puranam; phanish.puranam@insead.edu

Sanghyun Park and Phanish Puranam contributed equally to this work.

Received 20 September 2020; Revised 21 December 2020; Accepted 28 December 2020; Published 25 January 2021

Academic Editor: Wei Zhou

Copyright © 2021 Sanghyun Park and Phanish Puranam. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Learning by doing, a change in beliefs (and consequently behaviour) due to experience, is crucial to the adaptive behaviours of organizations as well as the individuals that inhabit them. In this review paper, we summarise different pathologies of learning noted in past literature using a common underlying mechanism based on self-confirming biased beliefs. These are inaccurate beliefs about the environment that are self-confirming because acting upon these beliefs prevents their falsification. We provide a formal definition for self-confirming biased beliefs as an attractor that can lock learning by doing systems into suboptimal actions and provide illustrations based on simulations. We then compare and distinguish self-confirming biased beliefs from other related theoretical constructs, including confirmation bias, self-fulfilling prophecies, and sticking points, and underscore that self-confirming biased beliefs underlie inefficient self-confirming equilibria and hot-stove effects. Lastly, we highlight two fundamental ways to escape self-confirming biased beliefs: taking actions inconsistent with beliefs (i.e., exploration) and getting information on unchosen actions (i.e., counterfactuals).

1. Introduction

The ability to learn is crucial to adaptive behaviour for agents in complex environments. Lacking omniscience, learning, a change in beliefs (and consequently behaviour) because of experience, is the primary mechanism through which an agent revises its beliefs to better represent the environment in which it finds itself and thus takes more adaptive actions. This is believed to be as true of individuals [1] as organizations [2] and other learning systems [3]. In particular, “learning by doing” characterizes many learning situations in organizations. It is a process through which agents learn from the results of their actions in a task environment (i.e., own experience). It is usually distinguished from social learning (i.e., learning from the experience of others) [4].

In learning by doing processes, two properties often cooccur. First, information about the environment is restricted to that resulting from actions taken by the agent, so-called “own-action dependence” [5] or endogenous

sampling [6]. In such situations, information that corresponds to unchosen actions is not available to the agent. Second, the agent is motivated to take actions that are likely to produce the best outcomes given current beliefs; agents act to “earn,” not only to “learn.” When these properties cooccur, the learning task is formally equivalent to the type of Markov decision problem known as a reinforcement learning problem [3].

For an example where both properties cooccur, imagine a situation involving hiring employees from three types of candidates (Table 1): A, B, and C. Employers are likely to choose employee type to maximize expected performance based on their beliefs (which may be incorrect to an unknown degree). As they interact with a chosen type of employee, they will gather information and update beliefs on that type. However, feedback on the unchosen types is not available for them, and their beliefs regarding those types will not be updated. This combination of own-action dependence and the agent’s selection of actions to maximize

TABLE 1: An example of SCBB in the context of hiring.

Employee types	True payoff	Employer's belief about payoffs	
		Type I	Type II
A	150	50	60
B	120	80	50
C	100	60	120

outcomes given current beliefs feature together in many learning by doing processes in organizations - whether in the context of manufacturing [7], service organizations [8], partner selection for alliances [9], or new product development [10].

In this analytical review paper, we describe self-confirming biased beliefs (SCBB) as a unified concept that forms the basis for understanding pathologies in “learning by doing” processes. SCBB are relevant whenever own-action dependence is present in learning contexts in which agents act to maximize expected returns given their beliefs. SCBB are biased in the sense that they are inaccurate representations of the environment, and they are self-confirming because acting upon these beliefs prevents their falsification [11]. In the example above, consider employers (type I) who believe that employee types A, B, and C yield 50, 80, and 60 units of payoff. Their true values are worth 150, 100, and 120 (i.e., the employer’s beliefs are biased). If employers take actions consistent with their beliefs, they will choose type B. The resulting outcome will be 100, thereby increasing their confidence in type B. However, they do not update their beliefs on A or C since they cannot observe their outcome (i.e., the counterfactual). Thus, type A or C will not be sampled even in the future, and this biased belief will perpetuate. SCBB are thus a particular type of attractor (i.e., stable fixed point) of learning by doing systems that can lock such systems into suboptimal actions [12].

Along with a formal definition of SCBB, we provide conceptual clarity by comparing SCBB with other related theoretical constructs across several literatures, including confirmation bias [13], self-fulfilling prophecy [14], self-confirming equilibria [15, 16], sticking points [17], and “hot-stove” effects [11]. In particular, we highlight that SCBB are a common concept underlying both inefficient self-confirming equilibria [15, 16] and hot-stove effects [11]. It can occur independently of confirmation bias or sticking points and act in opposition to self-fulfilling prophecies. This paper, thus, contributes to the literature on organizational learning by offering an integrative framework to understand the distinct nature of the pathologies associated with learning by doing, as well as a detailed analysis of one central concept, SCBB.

Last but not least, we elucidate two possible pathways to escape SCBB. The first involves forcing the agents to take actions inconsistent with their own beliefs, thus breaking the condition that agents maximize outcomes conditional on beliefs. In the previous example, employers making a decision inconsistent with their own beliefs (e.g., hiring type A while believing that type B is superior) may escape SCBB by correcting biased beliefs. This mechanism has been studied

extensively in terms of the exploration-exploitation trade-off in learning [18, 19]. The second solution, which is less widely understood, is to provide information on counterfactuals by escaping own-action dependence (i.e., information on unchosen actions). A modification of the task environment and agent behaviours that accomplish this is access to the experience of others. Social learning, even when there is no difference in the initial accuracy of beliefs across agents, can nonetheless break own-action dependence, if only to introduce noise to the focal agent’s beliefs by leveraging the diversity of erroneous beliefs [20, 21]. Again, in the hiring example, observing other employers (type II) who choose type C may reduce a focal employer’s confidence in the appropriateness of continuing with type B. This may eventually help them discover the correct belief (i.e., type A).

In the following section, we briefly review learning models in organization science. We then provide a formal definition of SCBB within the framework of a multiarmed bandit model followed by a comparison with related theoretical constructs. We also explore two mechanisms for escaping SCBB, exploration and social learning, and compare their viability in organizational contexts. Lastly, the implications of this study and notes on possible future extensions are provided.

2. Learning by Doing as a Form of Reinforcement Learning

Learning, revising beliefs based on available information, has been crucial in explaining many organizational phenomena [4]. In particular, there are two basic types of learning processes, learning by doing (equivalently learning from own experience) and social learning (or vicarious learning), learning from the experience of others. Of the two, learning by doing is the more fundamental process to understand since even social learning leverages the learning by doing of others. The centrality of learning by doing is also recognized as the principle of empiricism in the philosophy of science [22, 23].

Recent developments in machine learning have also highlighted other ways in which we might categorize learning problems. For instance, online learning describes a situation where information for learners unfolds over time and at a cost; the informational inputs to learning arrive in a staggered form. This is in contrast to offline learning, where the informational inputs are already present before the learning process begins (e.g., archival data) [3]. Learning by doing is, therefore, a form of online learning, but social learning may be either online or offline. Another categorisation that is prevalent in the machine learning literature is one that distinguishes supervised from unsupervised learning. In the former, the objective of what is to be learnt (i.e., an outcome to predict) is prespecified. For instance, an algorithm can learn how to predict creditworthiness based on past data on realized creditworthiness and applicant features. In the latter (unsupervised) form of learning, no prespecified dependent variable exists (e.g., clustering to find individuals who are demographically similar among voters).

Learning by doing almost always involves an objective in terms of performance and therefore can be seen as a form of supervised learning. Finally, when learners' choices determine both information generation process (i.e., own-action dependence) and their utilities, this constitutes a Markov decision problem known as a reinforcement learning task [3]. Learning by doing is, therefore, formally equivalent to a reinforcement learning task (which can also be described as both supervised and online).

While computer scientists are primarily interested in finding the optimal solution to learning problems, organization scientists have focused on the descriptive value of learning models. In particular, learning problems in organizations have often been described within the learning by doing framework and modelled using reinforcement learning tasks (e.g., [11, 24–29]; see [30] for a review). This is because organizational learning problems frequently meet the two conditions that define reinforcement learning problems.

First, choices in the learning process are often closely related to the effectiveness of an organization. As a consequence of pressures from competitors, stakeholders, or even colleagues, actions are usually motivated by the desire to obtain good outcomes given current beliefs. Second, in many organizational contexts, the value of alternatives can only be gauged by trying them (e.g., new product development, the adoption of organizational practices, or the choice of an alliance partner). The dynamic nature of organizational environments poses limits on offline learning since information generated in the past might not represent the current environment. In sum, learning by doing processes in organizations are well described in terms of reinforcement learning problems, in which subjective utility maximizers encounter a task environment with own-action dependence.

Next, we introduce the concept of self-confirming biased beliefs (SCBB) and how they may derail learning in reinforcement learning tasks (i.e., in organizational learning by doing processes).

3. Formal Definition

To provide a formal definition for SCBB, we describe the learning by doing process within the framework of a canonical reinforcement learning task, the multiarmed bandit problem. In this task, multiple alternatives exist, and an agent learns its values through repeated choices [3]. This model has been used extensively in organization science to analyse learning by doing processes, including individual-level processes [31], coupled learning process between individuals (or organizations) [24], and organization-level adaptation [25]. Along with a formal definition, we provide a numerical illustration of SCBB. We then illustrate the well-established result that exploration in choice can help escape SCBB. Finally, we introduce information on counterfactuals as a second mechanism that can also effectively combat SCBB.

Consider a task environment that consists of m possible alternative actions $\mathbf{A}: \{A_1, A_2, \dots, A_m\}$, and these map onto performance outcomes $\Pi: \{\Pi_1, \Pi_2, \dots, \Pi_m\}$. We assume that the alternative actions and their corresponding

outcomes are fixed and deterministic across time periods (i.e., a stable task environment). As the relationship $\mathbf{A} \rightarrow \Pi$ is unknown, an agent chooses an action based on its beliefs on possible alternatives— $\pi_t: \{\pi_{1,t}, \pi_{2,t}, \dots, \pi_{m,t}\}$. That is, the agent will choose an action that is believed to provide the greatest payoff at a given period (i.e., $\arg\max_{j \in \{1, \dots, m\}} \pi_{j,t}$), which is also called “greedy search.” Note that the agent’s belief on a specific action may not reflect its true value ($\Pi_i \neq \pi_{i,t}$). Also, beliefs at time t may differ from those at time t' for $t \neq t'$ as the agent updates its beliefs based on information gathered. When the agent takes a specific action, it will receive feedback for that action but not for other unchosen actions (i.e., own-action dependence). For simplicity, we assume here that there is no noise in feedback. That is, when the agent chooses action i , it will receive Π_i as the payoff in a deterministic manner. We consider SCBB in a noisy environment in Appendix A.

SCBB arise when an action with an incorrect belief is never sampled in the future. Formally, the condition under which the incorrect belief on the action i ($\Pi_i \neq \pi_{i,t}$) on the action i will be self-perpetuating (i.e., SCBB exists) is given by

$$\exists j \text{ that satisfies } \pi_{j,t} > \pi_{i,t} \text{ and } \Pi_j > \pi_{i,t}. \quad (1)$$

The agent will not sample the action i at time t since it believes that the action j is more attractive ($\pi_{j,t} > \pi_{i,t}$). Moreover, the true value of the action j is higher than a perceived payoff for the action i ($\Pi_j > \pi_{i,t}$). Thus, the agent will continue to believe that the action j is more attractive than the action i even when the agent learns the true value of the action j . Under this condition, the incorrect belief on the action i will never be falsified ($\Pi_i \neq \pi_{i,t}$, for $t' \geq t$).

Note that SCBB do not automatically imply poor performance. It is only when the incorrect belief about action i persists even though the action i is actually superior to the selected action j ($\Pi_i > \Pi_j$) that SCBB imply a learning pathology. Put differently, to earn more, the agent needs to correct SCBB on actions that are superior to the current one but not inferior ones. Thus, a learning system performs poorly because of SCBB when

$$\exists i, j \text{ that satisfy } \Pi_i > \Pi_j, \pi_{j,t} > \pi_{i,t}, \text{ and } \Pi_j > \pi_{i,t}. \quad (2)$$

3.1. An Illustration of Self-Confirming Biased Beliefs. To provide a numerical illustration of SCBB (code for reproducing the results from computational analysis for this paper is accessible via <https://github.com/sanghyunpark4/Self-confirming-biased-beliefs/blob/main/SCBB.py>), we model a task environment as a multiarmed bandit task with 50 alternatives ($m = 50$), with their corresponding performance outcomes (Π_i where $i \in \{1, 2, \dots, 50\}$) drawn from the uniform distribution in $(0, 1)$. For an agent learning in this environment, we assume that it possesses its own belief for each alternative at the initial stage of learning (i.e., its prior), which is also drawn from the uniform distribution in $(0, 1)$. In other words, the agent starts with unbiased prior in terms of the distribution. Lastly, at each point in time, the agent

chooses the alternative that is believed to offer the greatest payoff (maximizes subjective expected utility) and updates its belief by following a Bayesian norm for updating (i.e., averaging past payoffs). We demonstrate below that the own-action dependence condition is necessary and sufficient for an adaptive agent who acts as above to be susceptible to SCBB.

The pattern of SCBB in the learning by doing process is robust to other model specifications, the number of alternatives, the distribution of payoffs and priors, and the updating rule (Appendix B). The model parameters are summarized in Table 2. As our model has stochastic components (i.e., payoff and prior distribution), all data points in the following figures were averaged over 10,000 repeated simulations to reduce statistical errors (we choose the sample size by setting a tolerance level at 5% for the proportion of the best choice at the steady-state. To be specific, we generate 10,000 samples for each sample size (i.e., 10, 100, 1,000, 10,000, and 100,000) and check whether the range of proportions of the best choice is smaller than 5%. We find that the pattern of SCBB is robust regardless of the sample size (Appendix C)).

Figure 1 illustrates SCBB for different information conditions. First, our result shows that incorrect beliefs (measured as the Manhattan distance between belief and reality vectors) in the system with own-action dependency persist, while they eventually disappear if either complete information on the consequences of taking all actions or even on a randomly selected action is provided to the agent (see Figure 1(a)). Second, under own-action dependence, learning by doing produces lock-in because of SCBB. In Figure 1(b), only about 14% of cases among 10,000 repeated simulations reach the global optimum. Interestingly, the system with random information does not suffer as much from SCBB, even though the information given to the agent is incomplete. The system can still reach the best alternative even though it takes a longer time compared to that with complete information. In other words, own-action dependence (combined with the agent's actions that maximize expected payoff conditional on beliefs) is the root cause of SCBB rather than the amount of information per se.

Lastly, SCBB may not necessarily lead to inferior short-term performance. In particular, in our illustration, the system with own-action dependence outperforms that with random information until $t \leq 730$ (see Figure 1(c)). In contrast, the probability of choosing the optimal alternative under random information exceeds that under own-action dependence around $t = 100$. The trade-off is between SCBB producing premature convergence to a good but not optimal action, whereas random information provision produces an opportunity cost (in terms of not knowing outcomes for action chosen) that may only be offset given time [19].

4. Differentiating SCBB from Related Constructs

SCBB are distinct from confirmation bias, which refers to “*the seeking or interpreting of evidence in ways that are partial to*

existing beliefs, expectations, or a hypothesis in hand [13].” This is a cognitive bias in information processing driven by reliance on heuristics or avoidance of cognitive dissonance [32]. However, the root causes for SCBB are a task environment that forces endogenous sampling and agents who maximize returns conditional on beliefs; the agent may process the resulting information without any biases of the form noted above and still succumb to SCBB. As we demonstrated above and as is well recognized, SCBB can arise even when agents begin with unbiased priors and follow Bayesian norms for updating [6, 33].

SCBB also differ from the self-fulfilling prophecy, which refers to “*a false definition of the situation evoking a new behavior which makes the originally false conception come true* [14].” Its underlying mechanism is that the task environment is responsive to behaviors in a way that reduces bias in beliefs. For example, teachers' expectations of students can be self-fulfilling because students react to teachers' behaviors induced by their expectations [34]. In other words, a self-fulfilling prophecy illustrates a process in which biased beliefs become correct representations of reality because of changes to the task environment caused by the agent's actions. On the contrary, SCBB describe the persistence of biased beliefs despite the learning process. In fact, it can be shown that a responsive task environment, a necessary condition for the self-fulfilling prophecy, will reduce SCBB (Appendix D).

To further distinguish SCBB from other related constructs, it is useful to note that there can ultimately be only two sources of the beliefs that produce SCBB: erroneous priors and noisy feedback. For instance, when the agent believes that a particular alternative is unattractive at the initial stage, it will not be sampled. Thus, even when such a belief is incorrect (i.e., a false-negative belief), it will not be revised. Further, even when the agent has sampled the optimal alternative (i.e., the highest expected payoff), it may deviate from that alternative in the subsequent periods if the realized payoff is below the expected payoff due to noisy feedback (i.e., “hot-stove” effect [11]). In n -agent games, players who are subject to SCBB may end up in suboptimal self-confirming equilibria (there is a possibility that incorrect beliefs at off-path information sets may persist), which diverge from Nash equilibria [15, 16]. Thus, SCBB are a superset of both inefficient self-confirming equilibria (because they can exist even with a single agent) and “hot-stove effects” (because they can exist even when there is no noise in payoffs).

Lastly, SCBB are also distinct from sticking points, which have been defined in the context of local search on rugged landscapes. These refer to “*a configuration of choices such that once the firm arrives at the configuration, the firm will never deviate from it* [17].” While both SCBB and sticking points are attractors (i.e., stable fixed points) of the adaptive system, the source of stability varies. On the one hand, the interdependency between elements of the system is a root cause of sticking points. An accurate assessment of a configuration combined with a local search constraint produces fixation for the system in the case of sticking points. On the other hand, the own-action dependency combined with the

TABLE 2: Model parameters.

Parameter	Definition	Possible values	Sampled space
m	The number of alternatives	Positive integer	50
T	Learning period	Positive integer	1~100,000
τ	Degree of exploration	Real number (0, ∞)	Baseline case: 0 exploration: 0.01~1
N	The number of agents	Positive integer	Baseline case: 1 social learning: 2 larger systems: 1~15
γ	The degree of responsiveness	Real number ($-\infty$, ∞)	Investment task: 0.1 depletion task: -0.1

FIGURE 1: Self-confirming biased beliefs for different information conditions. (a) Degree of biased beliefs (i.e., the distance between belief and reality), (b) probability of choosing the best alternative, and (c) cumulative performance (divided by time period).

tendency to maximize payoffs based on beliefs causes SCBB. Thus, while sticking points and SCBB are both instances of the interactions between task environments and agent properties (i.e., Herbert Simon’s famous “scissors” [35]), they are also qualitatively different. Specifically, in SCBB, the agent’s beliefs must be biased in a way such that acting upon the beliefs prevents the generation of evidence that may falsify the incorrect belief. Thus, SCBB can emerge even without interdependency in the task environment (e.g., as in

the previous illustration) or a local search constraint, both of which are necessary for sticking points.

5. How to Escape Self-Confirming Biased Beliefs

Given that these are the necessary conditions, disrupting the agent’s tendency to maximize payoffs based on their beliefs or breaking up own-action dependency is the only possible path for escaping SCBB. The first path involves forcing the

agent to engage in “exploration,” which is defined as taking actions inconsistent with current beliefs [36]. In learning models, the exploration process has been extensively studied [18, 19]. By sampling actions that would not be chosen under the existing belief system (i.e., taking those actions believed to be less attractive), the agent may deviate from SCBB. In the previous hiring example in Table 1, employers can escape SCBB by choosing type A, which is inconsistent with their beliefs, and correct biased belief. This reveals the well-known benefit of exploration in learning by doing, and it is common to introduce random noise into action selection stages in learning models (e.g., ϵ -greedy, Luce’s choice rule, softmax [3], or maximum entropy [37]). Yet, implementing this is by no means an easy injunction for human actors to follow, as demands for consistency, justification, and explanation of actions are usually quite high in social settings. This prompted James March to memorably call for “technologies of foolishness” that would enable agents to take actions inconsistent with current best beliefs [36].

In organizational contexts, exploration includes experimentation, search, innovation, and variation, which contradict a tendency to behave consistently, synonymous with exploitation (e.g., refinement, efficiency, productivity, and variance reduction). In overcoming the tension between exploration and consistency, either relaxing demands for consistency or separating explorative activities to a different organizational unit is often cited as feasible policies [38]. For example, an organizational culture that values both innovation and efficiency may allow individuals to engage in innovative activities without damaging quality or efficiency [39]. Alternately, the tension can also be resolved by isolating exploration activities from exploitative activities. The separation can be achieved at three different levels: organizational separation (e.g., having an R&D department), temporal separation (i.e., sequential between exploration and exploitation), or domain separation (i.e., exploring in some domains while exploiting in other domains) (see [38] for a review). However, organizational scholars commonly agree that maintaining a sufficient degree of exploration is a demanding task in organizational contexts [19].

A third, less remarked upon approach is to provide evidence to the agent that might be independent of its own actions (i.e., supply information on counterfactuals). Figure 1 shows that providing information for a randomly selected alternative (instead of the actual action taken) can resolve SCBB. This might be hard to practically implement in most task environments in which learning by doing occurs. However, a possibility is to exploit the fact that the experiences of others can be a source of information on counterfactuals [21]. To illustrate this mechanism, consider that the employers in the previous example can alter their own beliefs when they observe other employers (type II) who believe (also erroneously) that type C is more attractive than type A or B; but if this social learning reduces their confidence in the appropriateness of continuing with type B, this may eventually help them discover the optimal action (i.e., type A).

The ability of social learning to produce counterfactual information will, of course, depend on how different the

copier and the copied are. Therefore, as long as the agents in the same task environment take different actions either because of different priors, because they obtained differences in feedback from the same action (e.g., through noisy payoffs), or differences in how they learn from feedback, copying each other can be a mechanism to break own-action dependence. Employers (type I) can correct SCBB on type C by gathering information on that type from other employers (type II), which would not be available for isolated learners. The value of social learning, in this case, is not to transfer knowledge from the insightful to the ignorant, thus ratcheting up collective insight [40], but to escape from ignorance by exploiting diversity in the system (and hence its ability to generate counterfactual information).

We illustrate how exploration and counterfactuals from diverse others redress SCBB differently. To operationalize exploratory behaviour in terms of breaking the tendency to maximize based on beliefs, we assume that the agent follows the softmax rule [3]. To be specific, the probability that the agent chooses an action i in period t is given by

$$P_{i,t} = \frac{e^{(\pi_{i,t}/\tau)}}{\sum_{j=1}^m e^{(\pi_{j,t}/\tau)}}, \quad \text{where } \tau > 0. \quad (3)$$

Note that now all alternatives will be assigned a positive probability of being chosen. Thus, all alternatives will be sampled eventually (as $T \rightarrow \infty$), and the agent can escape SCBB by falsifying incorrect beliefs. The parameter τ represents the degree of exploration in the search process [3]. When τ is high, the selection of choices depends less on the subjective valuation of alternatives (i.e., more exploration). As $\tau \rightarrow 0$, the softmax rule converges to the greedy search rule in our baseline case. We make the exploration parameter endogenous to the received payoffs by assuming that $\tau_t = 1 - \text{Payoff}_{t-1}$ (we assume that when $\tau < 0.01$, agents follow a greedy search (i.e., choosing the best one in the belief system) to prevent division by zero). This assumption allows agents to stick to a good alternative once they found a satisfactory one, thereby isolating the effect of SCBB from that of constant exploration (which prevents exploitation of good choices once found) in understanding the propensity to choose the best alternative.

To illustrate social learning, we assume that there are two agents in the system without an ex-ante knowledge differential. They learn not only from their own experience but also from the other’s experience (i.e., action and corresponding payoffs). To be specific, we assume that they update beliefs by assigning equal weights on their own and the other’s experience, which implies that they are not biased in utilizing the information as the quality of information is independent of its source (the pattern of results that we illustrate here is robust to other specifications of exploration and social learning (i.e., the exploration parameter and weights on information source; see Appendices E and F)).

Figure 2 shows the degree of SCBB for exploration and social learning compared to a benchmark of providing

random information as in Figure 1. First, compared to the baseline case of own-action dependence, all three variations reduce biased beliefs; see Figure 2(a). In particular, all increase the probability of choosing the best alternative at the end of the learning period (Figure 2(b)). Second, the three interventions differ in their effectiveness in resolving SCBB to find the optimal alternative. Interestingly, we find that providing information on random action outperforms the other two mechanisms in the long run. This is because, in the other two mechanisms, the root cause (i.e., own-action dependence with agents who behave consistently with their beliefs) is only partially resolved. The explorative behaviour under the softmax rule is less prone to SCBB but cannot escape it entirely since exploitation is rarely zero.

For the system with social learning, belief systems of the agents converge over time through mutual imitation, thereby generating less counterfactual information. Social learning is, therefore, a self-limiting mechanism for escaping SCBB; its ability of producing this benefit declines with its application. At the same time, the intertemporal trade-off in redressing SCBB privileges social learning (Figure 2(c)). With information on a random action provided, the actions actually taken are not updated, losing out on the opportunity to benefit from finding good actions early on. Social learning does not have this problem while providing a useful source of counterfactuals. It helps to break own-action dependence early in the search process while, at the same time, allowing for exploiting good actions found early on (which neither exploration through softmax nor provision of information for randomly selected actions allow). Thus, not only is SCBB a fundamental cause of the exploitation and exploration trade-off [18, 19], but also social learning is a particularly effective means to optimize on this trade-off in a manner that breaks own-action dependence without sacrificing the gains from early successes, which other mechanisms like a constant level of nongreedy action selection do not provide.

This benefit of social learning can also be demonstrated in a larger system as long as agents hold heterogeneous beliefs and share counterfactual information. Figure 3 illustrates the impact of the system size (i.e., the number of agents) on SCBB when agents engage in social learning. In particular, our result shows that the probability of choosing the optimal action at the steady-state increases with the system size (Figure 3(a)). For example, while only 23% of cases reach the optimal action when the system consists of two agents, about 83% of systems find the optimal one when there are fifteen agents. As the system size increases, more diverse alternatives will be sampled unless agents start with identical priors (Figure 3(b)). Under identical priors, the multiagent system cannot enjoy the benefit of social learning in redressing SCBB. These results point to another form of the “wisdom of crowds” in remedying SCBB; as long as there is sufficient heterogeneity to produce counterfactuals during learning, the crowd can improve on the individual learner [41] (see also [42] for a similar result for problem-solving).

6. Discussion

In this review article, we summarise what we know about a pathology that is likely to arise in learning by doing systems, which can be traced to self-confirming biased beliefs (SCBB). In particular, we pinpoint two conditions that are jointly sufficient for learning systems to become susceptible to SCBB: own-action dependence and agents who take actions consistent with current beliefs (e.g., maximizing subjective expected utility). Under these conditions, adaptive agents may not be able to correct false-negative beliefs because acting consistently with those beliefs prevents actors from collecting information that would eliminate those beliefs. Thus, such incorrect beliefs can be self-perpetuating. Because the above two conditions are jointly sufficient to produce SCBB, the only way to escape SCBB is to either break own-action dependency (i.e., providing information on counterfactuals) or produce inconsistency in choice and belief (i.e., creating exploration) or both.

We provide a comparison between SCBB and related (or seemingly related) constructs in prior literature. On the one hand, SCBB are at the root of both suboptimal self-confirming equilibria [15, 16] and hot-stove effects [11]. On the other hand, SCBB differ from confirmation bias [13] and sticking points [17] and are potentially diminished by self-fulfilling prophecies [14]. SCBB can arise even without cognitive bias in information processing, interdependency of choices within the system, local search, or a responsive environment that adjusts to the agent’s actions.

We also reviewed different mechanisms that may help escape SCBB as well as their feasibilities. Although exploration helps a learning system escape SCBB, it is often demanding for individuals and organizations to engage in it. In addition to the natural tendency to behave consistently with own belief, social contexts (e.g., organizations) often require consistency and explanation of actions, which contradict exploration activities (e.g., experimentation, search, or variation). Further, despite SCBB, learners are likely to establish correct inference on the current best alternative since it has been sampled more than other underexplored alternatives. Ambiguity aversion would thus make exploration even more difficult, as an underexplored alternative, which is subject to a false-negative belief, is likely to be further discounted [16, 43]. The persistence of SCBB despite some exploration indicates that biased belief for the optimal alternative is deeply entrenched in the existing belief system.

Access to information on counterfactuals is an alternative mechanism to escape SCBB. In some contexts, this is easy to implement. For instance, investors in the capital markets can observe the performance of stocks that they did not invest [44], and employers might track the candidates that they did not hire (i.e., on LinkedIn). In other contexts, social learning (i.e., learning from others’ experience) can be a feasible solution that reconciles a demand for consistency in their private beliefs and actions with the ability of breaking own-action dependence. Under social learning, agents can benefit from gathering

FIGURE 2: Self-confirming biased beliefs for different correcting mechanisms. (a) Degree of biased beliefs (i.e., the distance between belief and reality), (b) probability of choosing the best alternative, and (c) cumulative performance (divided by time period).

FIGURE 3: The impact of the system size on self-confirming biased beliefs under social learning. (a) Probability of choosing the optimal alternative at the steady-state and (b) the number of alternatives sampled by the system.

information on counterfactuals even when each agent behaves consistently with their own beliefs. However, the nature of social influence is critical. For example, when individuals sample based on popularity (e.g., trying what the majority seem to be doing) without sharing experiences, they may develop “collective illusions” where beliefs are homogenized around popular but suboptimal alternatives [45].

Our discussion of SCBB has several implications for researchers interested in learning within and by organizations. The most basic point is that in learning by doing processes the amount of experience may not correspond to knowledge (i.e., the veridicality of beliefs) when actors have a strong incentive to earn, not (only) learn. An explorative agent with limited experience may have a better representation of the task environment than an exploitative agent with abundant experience (Figure 2(a)). Second, SCBB offer a distinct and parsimonious mechanism to explain persistent heterogeneity across organizations despite adaptive processes. In explaining the diversity of organizations (e.g., practices and forms), which is one of the central questions in organization science, previous approaches have relied on local search on a rugged fitness landscape [46] or rigidity (diminished sensitivity to feedback) of organizations combined with heterogeneous environments [47]. However, heterogeneity across organizations under homogenous environments even without any local search restrictions may persist due to SCBB. Organizations may lock into suboptimal practices not because they have ossified and do not learn or because their trajectories of local search have led them to a local peak, but because they maximize subjective expected utility; they do not see any reason to deviate from their current beliefs, which may, however, feature SCBB in their priors.

A natural extension of our work is to explore correcting mechanisms for SCBB in more detail, including their boundary conditions. Organization scholars have proposed several ways to balance the cost and benefit of exploration [38]. On the contrary, we have a limited understanding of the microprocesses through which agents learn from others’ experience and their boundary conditions for producing an accurate understanding of the task environment. Since social learning might be more feasible, imposing lower pressures of consistency or justification than exploration, in organizational settings, these questions are also practically relevant to improve learning by doing processes in organizations. The analysis of learning by doing and social learning (learning from others) may well benefit from a tighter integration, since even in the latter, as we have noted, ultimately one learns from the learning by doing of others.

Appendix

A. SCBB under Noisy Feedback

In the baseline model, we assume no noise in performance feedback (i.e., the performance feedback is deterministic). In

reality, performance feedback often entails noise [10, 26]. That is, under noisy feedback, performance feedback ($\widehat{\Pi}_{i,t}$) that the agent receives may deviate from the expected performance (Π_i). When performance feedback is stochastic, performance outcomes (Π) become the expected performance for each alternative ($\Pi_i = E(\widehat{\Pi}_i)$). Then, the condition for SCBB is given by

$$\exists j \text{ that satisfies } \pi_{j,t} > \pi_{i,t} \text{ and } \Pi_{j,LB} > \pi_{i,t}, \quad (\text{A.1})$$

where $\Pi_{j,LB}$ represents a lower bound of $\widehat{\Pi}_j$. Note that noise has two opposite impacts on SCBB. On the one hand, it reduces SCBB because $\Pi_{j,LB} \leq \Pi_j$. Put simply, noise in the performance feedback on the action j may allow the agent to correct biased beliefs on the action i when $\Pi_j > \pi_{i,t} > \Pi_{j,LB}$. On the other hand, noise in performance feedback is a source of incorrect beliefs ($\pi_{i,t} \neq \Pi_i$) along with erroneous priors, resulting in SCBB, “hot-stove effect [11].” The agent who sampled the best alternative may deviate from it in the subsequent periods when it receives unfavourable feedback due to noise. Thus, even when the agent starts with unbiased priors ($\pi_i = \Pi_i$ for $i \in \{1, \dots, m\}$), it is susceptible to SCBB when feedback is noisy.

B. Sensitivity Checks

The following figure describes sensitivity checks (Figure 4):

C. The Distribution of Estimates of the Proportion of the Best Alternative with regard to Sample Sizes

The following figure describes sample size (Figure 5):

D. SCBB in Systems with Responsive Environments

So far, we have described SCBB in the context of an agent learning by doing in an environment that is itself not agentic, which does not respond to the actions of the agents. However, there are learning situations in which the environment also reacts to the agent’s actions (e.g., the payoff to action may alter due to being selected). For instance, an agent’s payoff from an action may be enhanced through a repetitive selection of action (an “investment” task, such as A trusting B increases B’s trustworthiness [48] or an agent repeating a task increases the payoff for that task to the agent through some form of increasing returns [49]). Conversely, the selection of an alternative may diminish its payoffs (a “depletion” task, such as in the selection of a location to harvest or fish). To operationalize responsive environments, we assume that $\Pi_{i,t} = (1 + \gamma)^{n_{i,t}-1} \Pi_{i,0}$ where $n_{i,t}$ represents the number of times that the alternative i has been sampled during $t - 1$ periods and $\Pi_{i,0}$ is a payoff at $t = 0$. When $\gamma > 0$, the task environment is characterized as an investment task. On the other hand, $\gamma < 0$ represents a depletion task.

Our result shows that both forms of environmental reactance can mitigate the effects of SCBB. In the investment case, the action in which the agent is locked

FIGURE 4: (a) The number of alternatives (m), (b) updating rules (ERWA represents exponential recency weighted average where ϕ is a degree of recency. To be specific, the agent updates its belief on the action i as follows: $\pi_{i,t} = (1 - \phi)\pi_{i,t-1} + \phi\Pi_{i,t}$), and (c) distribution of payoffs.

objectively becomes the best over time so that there is no bias in terms of choosing the optimal alternative (Figure 6(c)), even though beliefs may still be biased

about other alternatives (Figure 6(a)). This result corresponds to a self-fulfilling prophecy [14]. In the depletion case, the selected action objectively becomes

FIGURE 5: Sample size.

FIGURE 6: (a) Degree of biased beliefs (i.e., the distance between belief and reality), (b) probability of switching action, and (c) probability of choosing the best alternative.

FIGURE 7: (a) Degree of biased beliefs (i.e., the distance between belief and reality); (b) probability of choosing the optimal alternative (note that when the high degree of exploration is exogenously given (e.g., $\tau = 0.1$), the agent may choose suboptimal alternatives not because of SCBB but because of exploration (i.e., noise in the choice process)).

FIGURE 8: (a) Degree of biased beliefs (i.e., the distance between belief and reality); (b) probability of choosing the optimal alternative (w_2 represents weight on information from the other, while w_1 represents weight on information for own action).

worse and also causes the agent to discard it (Figure 6(b)), eliminating bias by encouraging wandering over alternatives (Figure 6(a)).

E. Sensitivity Check for Exploration

The following figure describes sensitivity check for exploration (Figure 7):

F. Sensitivity Check for Social Learning

The following figure describes sensitivity check for social learning: (Figure 8)

Data Availability

The code for reproducing the results from computational analysis for this paper is made accessible to readers via GitHub

(<https://github.com/sanghyunpark4/Self-confirming-biased-beliefs/blob/main/SCBB.py>).

Disclosure

The ideas in this paper benefited from presentation at the James G March Memorial Conference held at Carnegie Mellon University in October 2019.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] I. Erev and A. E. Roth, "Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria," *American Economic Review*, vol. 88, pp. 848–881, 1998.
- [2] R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*, Prentice Hall, Englewood Cliffs, NJ, USA, 1963.
- [3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, USA, 2018.
- [4] L. Argote, *Organizational Learning: Creating, Retaining and Transferring Knowledge*, Springer Science & Business Media, Berlin, Germany, 2012.
- [5] P. Battigalli, A. Francetich, G. Lanzani, and M. Marinacci, "Learning and self-confirming long-run biases," *Journal of Economic Theory*, vol. 183, pp. 740–785, 2019.
- [6] G. Le Mens, J. Denrell, and J. Denrell, "Rational learning and information sampling: on the "naivety" assumption in sampling explanations of judgment biases," *Psychological Review*, vol. 118, no. 2, pp. 379–392, 2011.
- [7] L. Argote and D. Eppler, "Learning curves in manufacturing," *Science*, vol. 247, no. 4945, pp. 920–924, 1990.
- [8] E. D. Darr, L. Argote, and D. Eppler, "The acquisition, transfer, and depreciation of knowledge in service organizations: productivity in franchises," *Management Science*, vol. 41, no. 11, pp. 1750–1762, 1995.
- [9] M. P. Koza and A. Y. Lewin, "The co-evolution of strategic alliances," *Organization Science*, vol. 9, no. 3, pp. 255–264, 1998.
- [10] H. R. Greve, "Exploration and exploitation in product innovation," *Industrial and Corporate Change*, vol. 16, no. 5, pp. 945–975, 2007.
- [11] J. Denrell and J. G. March, "Adaptation as information restriction: the hot stove effect," *Organization Science*, vol. 12, no. 5, pp. 523–538, 2001.
- [12] S. H. Strogatz, *Nonlinear Dynamics and Chaos with Student Solutions Manual: with Applications to Physics, Biology, Chemistry, and Engineering*, CRC Press, Boca Raton, FL, USA, 2018.
- [13] R. S. Nickerson, "Confirmation bias: a ubiquitous phenomenon in many guises," *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998.
- [14] R. K. Merton, "The self-fulfilling prophecy," *The Antioch Review*, vol. 8, no. 2, pp. 193–210, 1948.
- [15] D. Fudenberg and D. K. Levine, "Self-confirming equilibrium," *Econometrica*, vol. 61, no. 3, pp. 523–545, 1993.
- [16] P. Battigalli, S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci, "Self-confirming equilibrium and model uncertainty," *American Economic Review*, vol. 105, no. 2, pp. 646–77, 2015.
- [17] J. W. Rivkin and N. Siggelkow, "Organizational sticking points on NK landscapes," *Complexity*, vol. 7, no. 5, pp. 31–43, 2002.
- [18] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, USA, 1975.
- [19] J. G. March, "Exploration and exploitation in organizational learning," *Organization Science*, vol. 2, no. 1, pp. 71–87, 1991.
- [20] Ö. Koçak, D. Levinthal, and P. Puranam, "The dual challenge of search and coordination for organizational adaptation: how structures of influence matter (working paper)," 2019.
- [21] S. Park and P. Puranam, "Learning what they think vs. learning what they do: the micro-foundations of vicarious learning," 2020, <https://arxiv.org/abs/2007.15264>.
- [22] J. Klein, E. N. Zalta, *Francis Bacon*, The Stanford Encyclopedia of Philosophy, Stanford, CA, USA, 2003, <https://plato.stanford.edu/entries/francis-bacon/>.
- [23] J. Hackett, E. N. Zalta, *Roger Bacon*, The Stanford Encyclopedia of Philosophy, 2007, <http://plato.stanford.edu/archives/sum2007/entries/roger-bacon>.
- [24] P. Puranam and M. Swamy, "How initial representations shape coupled learning processes," *Organization Science*, vol. 27, no. 2, pp. 323–335, 2016.
- [25] H. E. Posen and D. A. Levinthal, "Chasing a moving target: exploitation and exploration in dynamic environments," *Management Science*, vol. 58, no. 3, pp. 587–601, 2012.
- [26] T. Knudsen and K. Srikanth, "Coordinated exploration," *Administrative Science Quarterly*, vol. 59, no. 3, pp. 409–441, 2014.
- [27] O. Baumann, "Models of complex adaptive systems in strategy and organization research," *Mind & Society*, vol. 14, no. 2, pp. 169–183, 2015.
- [28] E. Lee and P. Puranam, "The implementation imperative: why one should implement even imperfect strategies perfectly," *Strategic Management Journal*, vol. 37, no. 8, pp. 1529–1546, 2016.
- [29] N. Stieglitz, T. Knudsen, and M. C. Becker, "Adaptation and inertia in dynamic environments," *Strategic Management Journal*, vol. 37, no. 9, pp. 1854–1864, 2016.
- [30] P. Puranam, N. Stieglitz, M. Osman, and M. M. Pillutla, "Modelling bounded rationality in organizations: progress and prospects," *Academy of Management Annals*, vol. 9, no. 1, pp. 337–392, 2015.
- [31] J. S. Chen, D. C. Croson, D. W. Elfenbein, and H. E. Posen, "The impact of learning and overconfidence on entrepreneurial entry and exit," *Organization Science*, vol. 29, no. 6, pp. 989–1009, 2018.
- [32] R. J. MacCoun, "Biases in the interpretation and use of research results," *Annual Review of Psychology*, vol. 49, no. 1, pp. 259–287, 1998.
- [33] J. Denrell, "Why most people disapprove of me: experience sampling in impression formation," *Psychological Review*, vol. 112, no. 4, pp. 951–978, 2005.
- [34] J. E. Brophy, "Research on the self-fulfilling prophecy and teacher expectations," *Journal of Educational Psychology*, vol. 75, no. 5, p. 631, 1983.
- [35] H. A. Simon, "Invariants of human behavior," *Annual Review of Psychology*, vol. 41, no. 1, pp. 1–20, 1990.
- [36] J. G. March and J. P. Olsen, *The Technology of Foolishness*, Universitetsforlaget, Oslo, Norway, 1976.
- [37] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," *AAAI*, vol. 8, pp. 1433–1438, 2008.

- [38] D. Lavie, U. Stettner, and M. L. Tushman, "Exploration and exploitation within and across organizations," *Academy of Management Annals*, vol. 4, no. 1, pp. 109–155, 2010.
- [39] E. Miron, M. Erez, and E. Naveh, "Do personal characteristics and cultural values that promote innovation, quality, and efficiency compete or complement each other?" *Journal of Organizational Behavior*, vol. 25, no. 2, pp. 175–199, 2004.
- [40] J. Henrich, *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*, Princeton University Press, Princeton, NJ, USA, 2017.
- [41] H. Piezunka, V. A. Aggarwal, and H. E. Posen, "Learning-by-participating: the dual role of structure in aggregating information and shaping learning," INSEAD, Fontainebleau, France, Working paper, 2020.
- [42] L. Hong and S. E. Page, "Problem solving by heterogeneous agents," *Journal of Economic Theory*, vol. 97, no. 1, pp. 123–163, 2001.
- [43] P. Battigalli, E. Catonini, G. Lanzani, and M. Marinacci, "Ambiguity attitudes and self-confirming equilibrium in sequential games," *Games and Economic Behavior*, vol. 115, pp. 1–29, 2019.
- [44] F. A. Csaszar, "Organizational structure as a determinant of performance: evidence from mutual funds," *Strategic Management Journal*, vol. 33, no. 6, pp. 611–632, 2012.
- [45] J. Denrell and G. Le Mens, "Information sampling, belief synchronization, and collective illusions," *Management Science*, vol. 63, no. 2, pp. 528–547, 2017.
- [46] D. A. Levinthal, "Adaptation on rugged landscapes," *Management Science*, vol. 43, no. 7, pp. 934–950, 1997.
- [47] M. T. Hannan and J. Freeman, "The population ecology of organizations," *American Journal of Sociology*, vol. 82, no. 5, pp. 929–964, 1977.
- [48] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," *Games and Economic Behavior*, vol. 10, no. 1, pp. 122–142, 1995.
- [49] B. Levitt and J. G. March, "Organizational learning," *Annual Review of Sociology*, vol. 14, no. 1, pp. 319–338, 1988.