AI Adoption and System-Wide Change

Ajay Agrawal, Joshua S. Gans, Avi Goldfarb

May 2021

Abstract

Analyses of AI adoption focus on its adoption at the individual task level. What has received significantly less attention is how AI adoption is shaped by the fact that organisations are composed of many interacting tasks. AI adoption may, therefore, require system-wide change which is both a constraint and an opportunity. We provide the first formal analysis where multiple tasks may be part of a modular or non-modular system. We find that reliance on AI, a prediction tool, increases decision variation which, in turn, raises challenges if decisions across the organisation interact. Modularity, which leads to task independence rather than system-level inter-dependencies, softens that impact. Thus, modularity can facilitate AI adoption. However, it does this at the expense of synergies. By contrast, when there are mechanisms for inter-decision coordination, AI adoption is enhanced when there is a non-modular environment. Consequently, we show that there are important cases where AI adoption will be enhanced when it can be adopted beyond tasks but as part of a designed organisational system.

Keywords: artificial intelligence, machine learning, modularity, systems

1   Introduction

Over the past decade, artificial intelligence (AI) has emerged as a potential general purpose technology (Cockburn et al. (2019)). Spurred on by advances in machine learning, the cost of prediction across various domains has started to fall at an accelerating pace (Agrawal et al. (2018a)). This raises interesting questions of where AI will be adopted and also its potential disruptive impact on employment and businesses (Gans and Leigh (2019), Frey and Osborne (2017), Brynjolfsson and McAfee (2017)).

To date, our conception of AI adoption has mainly operated at the level of a task or decision (e.g., Frank et al. (2019); Acemoglu and Restrepo (2018)). To forecast the potential impact of AI on employment, for example, there have been numerous exercises designed to identify jobs at risk from AI, the tasks that comprise jobs that are at risk, and the more general impact of automation on the workplace (Webb (2020); Brynjolfsson and Mitchell (2017); Brynjolfsson et al. (2018); Felten et al. (2018)). That said, some have questioned whether this task-level focus is appropriate. Bresnahan (2020) argues that AI is an information technology and traditionally such technologies have required organisational redesign to be fully adopted. This is readily apparent in patterns of adoption of earlier generations of IT (Bresnahan and Greenstein (1996); Bresnahan et al. (2002); Aral et al. (2012); Dranove et al. (2014)).

Bresnahan (2020) challenges the idea that AI adoption can be analysed at the task level independent of the organisational context the task lies in. Bresnahan identifies the degree of modularity in the organisation as a predictor of AI adoption. When an organisation is non-modular, changing the nature of decision-making in one part – as would arise with AI adoption – can require changes in decision-making and practices elsewhere. The need for adjustment throughout a non-modular organisation can, he argues, cause a barrier to the adoption of AI. Instead, Bresnahan forecasts that AI will be primarily adopted in existing modular organisations and might only be adopted in other organisations should those organisations be redesigned to be modular.

We build a model of the role of modularity in AI adoption. We consider a firm where value arises from the outcomes of two decisions (akin to tasks). Modularity is the degree to which the firm receives a payoff from one decision even if the two decisions are not properly aligned. In the baseline model, the decision-makers do not know the external state and hence do not know the correct action. Therefore, they select the action with the highest likelihood of being correct. There is no reason for the decision-makers to communicate, irrespective of the degree of modularity. They both do the same action at all times and are often, but not always, correct.

AI helps decision-making by providing information on the correct choice for one of the decisions. In this respect, AI adoption itself occurs with respect to a single task. With the AI, there can be a benefit to deviating from the focal action. However, if the organisational structure is unchanged with no improvements to intra-organisational communication, we find that AI adoption is more valuable in modular organisations. This result is consistent with Bresnahan's intuition: It is easier to adopt AI in modular organisations because modularity means less need to coordinate. One decision-maker can respond to incoming information without affecting the outcomes for the other decision-maker.

We next consider communication as a type of organisational change.[1] With communication, AI is especially valuable to non-modular organisations that can take advantage of the ability to get the first decision right (because of the AI) and then coordinate the second decision. With communication, AI will benefit non-modular organizations and these organisations will benefit even more than modular organisations, with or without communication.

We see this framework as informing when to focus on AI adoption with a task versus a systems lens. It is appropriate to focus on AI adoption at the task level when examining modular organisations that do not change. In contrast, a systems focus is appropriate when the organisation can change how it operates, for example by facilitating communication and hence coordination. In this case, AI adoption cannot be analysed without considering the potential for systems-level change to enable non-modular organisations to succeed.

Throughout, we provide examples from the retail industry, emphasizing how AI is already enabling modular organisations and how, with better communication, it could also enable non-modular organisations.

2 Model Set-Up

The model here is inspired by Van den Steen (2017), although it addresses a distinct research question.[2] Suppose that an organisation's return, R, depends on the outcomes of two decisions, $\{D_1, D_2\}$, indexed by k.[3] Each decision results in the choice of an action, $a_k \in A_k$,

---

[1] One interpretation of investment in communication within an organisation is that this is a co-invention that enables AI; see Bresnahan and Greenstein (1996); Bresnahan et al. (2002); Aral et al. (2012); and Dranove et al. (2014).

[2] Van den Steen (2017) examines the role of strategic leadership in coordinating actions within an organisation and seeks to identify the set of decisions that a leader will focus on. The present paper does not examine such strategy at all and instead focuses upon a question of technology adoption and the organisational developments that may complement it. Thus, the model here is more specific in some dimensions and more general in others than Van den Steen's.

[3] Van den Steen (2017) considers a model that can comprise more than two decisions. While that is possible for the model considered here, this dimension proves unnecessary for any qualitative conclusions drawn below and so we chose to focus on just two decisions.

a set with M_k elements. D_1 is a *focal* decision whose "correctness" depends upon matching an action with an external state. A stand-alone correct D_1 results in an increment to return of α compared with an incorrect decision. D_2 is a "supporting" decision that creates value when aligned with the focal decision. If two decisions, $\{D_1, D_2\}$, are correctly aligned, then this results in an incremental return of γ relative to the case where those decisions are incorrectly aligned (see Figure 1).[4] Throughout, we consider the example of a retailer, in which the decisions are what to stock in inventory and what to recommend to the customer.

The "correctness" of a decision is modelled here in a reduced form way. We suppose that the correct stand-alone decision is associated with a state, $T_1 \in A_1$, knowledge of which reveals the correct decision. This state could be driven by an assessment of the external environment for a decision and/or an agent's judgment regarding the trade-offs and risks associated with particular actions. For instance, a retail manager may be considering a decision of how much to stock based on a prediction of future demand as well as the relative costs associated with errors in that forecast (inventory holding costs versus lost sales due to stock outs). Based on that prediction, at a given time, there is an assessment of the state and associated optimal action. If that state is correctly identified and the associated action taken, there is a boost to the organisation's return of α. If not, there is no such boost.[5]

We treat the correctness of the supporting decision similarly. We suppose that whether $\{D_1, D_2\}$ are aligned is associated with a state $T_{12} \in \{A_1, A_2\}$. If $\{a_1, a_2\} = T_{12}$, then the decision is "correct" and it contributes γ to project return. Otherwise, there is no contribution. We assume that T_{12} is a bijection (i.e., a one-to-one correspondence) where for every a_1 there exists an action a_2 that creates alignment. Thus, so long as this relationship is known, there is an alignment incentive to choose the actions that selected that state for $\{D_1, D_2\}$. It is assumed that $\alpha, \gamma > 0$.

The stand-alone and alignment contributions, α and γ, can be realised if D_1 and D_2 are chosen correctly. We assume that if *both* D_1 and D_2 are correct – that is, $a_1 \in T_1$ and $\{a_1, a_2\} \in T_{12}$ – a contribution of $\beta > \alpha + \gamma$ can be generated. We will call β a *synergistic* contribution. As we will note shortly, this particular opportunity defines both the benefit of non-modular systems as well as driving the difficulty in managing non-modular systems.

---

[4]In Van den Steen (2017), the alignment value can be different depending upon which decision-maker is doing the alignment. Here, however, we will assume that alignment is the responsibility of the agent who has authority over D_2. This is a special case of Van den Steen (2017) when the alignment value arising from D_1 is very small or the agent responsible for D_1 has no knowledge of the action chosen for D_2 or knowledge of what the correct alignment action would be.

[5]In many decisions, the "distance" from the optimal or correct decision matters. Here we abstract from those considerations, but it could be imagined that α_k is a measure of the loss from deviating from the optimum, and a tolerance for errors would be reflected in a relatively low α_k.

Figure 1: **Decisions and Contributions**
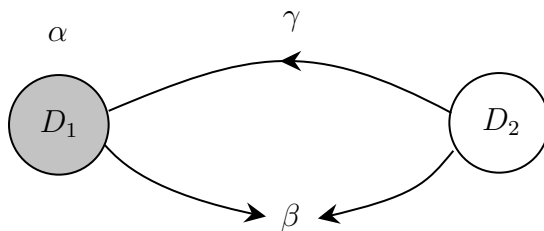
2.1 Modularity

What are the characteristics of an organisation that may impact on AI adoption? Bresnahan (2020) argues that a modular organisation will face fewer barriers to adopting AI than one that is not modular. Modularity does not have a commonly agreed upon definition (Campagnolo and Camuffo (2010)). In the study of organisations, its origins are most associated with Simon (1962) and then developed by Baldwin and Clark (2000). In a modular system, actions taken in one component of the system do not impact on those taken in other components and vice versa. At an organisational level, modularity relates to how loosely different parts of the organisation are coupled together (Schilling (2000)).

As a general matter, the main trade-off between a modular and a non-modular organisation is that the former is easier to manage than the latter; it does this by sacrificing certain opportunities for creating synergistic value. However, it is also the case that modularity makes it easier for decisions to create value by being aligned. In our case, an alignment contribution of γ captures this possibility, while the synergistic contribution of β captures the additional challenge of coordination when the focal decision may itself be varying. In this sense, the degree of modularity is a measure of the weight placed on achieving α and γ rather than β.[6]

We choose a particular functional form for the organization's return, R, that allows us to clearly capture these trade-offs. Specifically, we assume that:

$$R = m(\alpha I_{a_1=T_1} + \gamma I_{\{a_1,a_2\}=T_{12}}) + (1-m)\beta I_{a_1=T_1} I_{\{a_1,a_2\}=T_{12}}$$

Here $m \in [0,1]$ is the *degree of modularity* in the organisation. Note that as m rises, more weight is placed on being able to realise value from D_1 and D_2 independently of the correctness or performance of other decisions. As Baldwin and Clark (2000) note, a modular organisation simplifies the level of interactions and communication needed to align different

---

[6]In the appendix, we consider an amended model where agent 2 can choose whether to align with D_1 or not. This allows us to consider modularity by *act* (rather than just modularity by *design*, which operates through m). We show that all of the conclusions regarding modularity derived below carry through to this more expansive treatment of modularity.

components or reduce conflicts between them by being tolerant for errors made elsewhere. By contrast, absent these factors, decisions might be tightly coupled and overall performance becomes more sensitive to mistakes made in one area.[7]

The definition of modularity here links organisational performance to the joint correctness of the decisions in the organisation. It is useful to interpret this in the context of an example. For instance, Bresnahan (2020) argued that Amazon had a modularised online store. In particular, if Amazon's recommendation engine changed how it recommended products to different customers, this "would not require changing other elements extensively because of the modularisation." He does not outline these other elements, but one could imagine elements such as what products to stock, how to distribute those products to customers, and so on. He attributes this to things that Amazon already put in place that allowed it to scale its operations. Note that this does not mean that other parts of Amazon's organisation do not have to be aligned with what it is recommending. Indeed, it makes little sense to recommend products to customers that Amazon does not stock. Thus, a stocking decision is arguably correct so long as what is being recommended is in stock. This enables Amazon to generate a contribution such as γ.

But what happens if Amazon recommends a product to a customer that isn't the best match for that customer? In this case, is there any value to having the recommended item in stock? Bresnahan argues that there is because the convenience benefits of Amazon means that a customer may well order a product that isn't its best match if it is in stock. Thus, α may not be realised (because the customer is not as satisfied as they might be) but γ is generated if the recommended item is in stock. Thus, how well D_1 (the recommendation decision) performs does not impact, in Bresnahan's telling, the value that can be realised by an alignment decision (D_2).

By contrast, what if consumers would not order a product unless it is their best match? In that situation, having the recommended item in stock is not enough to generate an order and some contribution to the organisation. Instead, it is important that the recommended item be the best match to generate an order and realise the value of having the item in stock. When the performance of D_1 impacts on the performance of D_2, we say that the organisation is not modular. In effect, the operation of the online store cannot have high performance independently of the performance of the recommendation engine. Only when high performance happens jointly is the contribution (in this case, β) realised. This happens, for example, for an online service such as StitchFix, which sends customers clothing items each month. If they were to send clothing items that the customer does not want and thus

---

[7]The baseline structure in Van den Steen (2017) assumes that $m = 1$ and does not consider situations where value requires *both* a decision to be correct and other decisions to be correctly aligned.

returns, then the fact that those items were procured by StitchFix and shipped in the first place does not generate value. Perform well and the customer keeps the product without the inconvenience of shopping (generating value of β rather than $\alpha + \gamma$ that would be earned in a store where the customer has to shop).

2.2 Information Structure

Van den Steen (2017), assumes that each of T_1 and T_{12} are equally likely and drawn from an infinite set, with $M_k \to \infty$ for all k. Thus, there is a zero probability that one agent could guess the "correct" action in the absence of knowledge of the relevant state. We retain that assumption here.[8]

There are two agents in the model, each responsible for D_1 and D_2 respectively, with their main distinction being what they can observe rather than differing objectives (i.e., each shares the objective of maximising R). We will assume that agent 1, who "owns" a decision, D_1, is the only agent who can potentially observe, in the absence of communication, the outcome T_1. Thus, agent 2 does not know 1's correct stand-alone decision. With respect to alignment, we assume that only agent 2 can know T_{12} and they possess this knowledge for any given a_1.[9] Agent 2, however, does not directly observe a_1. Thus, achieving alignment requires agent 2 to correctly infer or predict agent 1's choice so as to select the a_2 that will align with that choice from T_{12}. Recall, that T_{12} is a bijection. Thus, there is an a_2 that will generate a contribution if aligned with the correct a_1.

Finally, it is assumed that for D_1 there is a focal action; that is, there is one action, a_S, that is included in T_1 with probability $\rho(> \frac{1}{M_1-1})$, while the other actions are in T_1 with probability $(1-\rho)\frac{1}{M_1-1} \to 0$ as $M_1 \to \infty$. Thus, in the absence of information, agent 1 can always generate a contribution of at least $m\rho\alpha$ by choosing the focal action, a_S. For example, if a retailer does not know who is shopping, they are likely to recommend their most popular product to everyone.

2.3 Equilibrium with no prediction

AI adoption will involve a prediction of T_1 and agent 1 following that prediction (Agrawal et al. (2018b)). To begin we consider the outcome without AI; i.e., the (Nash) equilibrium

---

[8]If each M_k is finite, then this introduces the possibility of alignment by "chance." While that may meaningfully reflect a more simple environment, it is strategically uninteresting, and not accounting for this would not change the qualitative results below.

[9]It is possible that agent 2 does not have this knowledge, which introduces a challenge in aligning. However, accounting for the possibility that 2 only knows T_{12} with some probability adds complexity without altering the qualitative results below.

outcome when agent 1 does not know T_1. In this case, the fact that D_1 has a focal action, a_S, that is the correct stand-alone action more frequently than any other makes this a natural choice to focus on in considering equilibrium outcomes. We assume that both agents act to maximise R.[10]

The following proposition characterises a Nash equilibrium outcome.

Proposition 1 *There exists a Nash equilibrium where $a_1 = a_S$ always with equilibrium payoffs*

$$R = m(\rho\alpha + \gamma) + (1 - m)\rho\beta$$

Proof. We begin by assuming that agent 1 will always choose a_S in the absence of information regarding T_1. (We will establish below that they choose this in equilibrium). Thus, $E[I_{a_1=T_1}] = \rho$. Agent 2 will use this information, choosing their action to align with D_1. Specifically, as agent 2 knows T_{12}, then $E[I_{\{a_1,a_2\}=T_{12}}] = 1$. Given this, the expected payoff is as stated in the proposition.

Can agent 1 improve this payoff by selecting another action, $a_1 \neq a_S$? In this case, if agent 2 is expecting $a_1 = a_S$, they will not align with D_1 and so $E[I_{a_1=T_1}] = 0$. Thus, R falls to $m\rho\alpha$. Thus, in this case, the conjectured equilibrium is an actual equilibrium. ∎

Proposition 1 shows that, in the focal point equilibrium (where $a_1 = a_S$ always), agent 2 is able to align with that action. Modularity (m) impacts on this payoff. Note that the equilibrium payoff without prediction is increasing in m if:

$$\rho\alpha + \gamma > \rho\beta \implies \frac{\gamma}{\beta - \alpha} > \rho \tag{1}$$

If ρ is low, this means that the loss from there being no ability of agent 1 to forecast the correct state of the world is high. In this situation, an organisation that is more modular generates a higher payoff as the challenges to realise synergies are correspondingly higher. This captures the trade-off noted earlier that modularity makes it easier to realise the value of decisions when coordinating their performance is challenging due to uncertainty.

The equilibrium in Proposition 1 is not, however, the only equilibrium outcome. If agent 1 chooses, with certainty, any action, then agent 2 will optimally align to that action. Agent 1's expected stand-alone contribution will be lower. Thus, all other equilibrium outcomes result in a lower R. For that reason, we will focus on the "focal point" equilibrium outcome where agent 1 chooses the action most likely to be the correct action.

---

[10]An alternative outcome would be that they maximised the own contribution to R. We focus here instead on the team outcome to avoid introducing other incentive considerations.

It is important to emphasise here that having a focal action that is consistently chosen by agent 1 allows for alignment to occur *without any explicit coordinating mechanism such as communication or centralised direction.* This will be important because if agent 1 has a prediction for the correct action and chooses to act on that, this will disrupt the ability of agent 2 to align with that choice and will enhance the value of having an explicit coordinating mechanism.

3   The AI Adoption Decision

We now turn to consider how this equilibrium outcome changes whether the organisation should adopt AI so that agent 1 receives a signal or prediction of T_1. First, we examine how the choices and performance of the organisation change as a result of being able to receive that prediction. Answering this question is a precursor to examining whether the organisation will choose to adopt AI or not.

We assume that there exists a perfect prediction of T_1, i.e., a signal that reveals which action is the correct stand-alone action for D_1.[11] In this case, should agent 1 choose the signaled action, the stand-alone contribution will be α. Thus, compared with no information, a perfect prediction allows agent 1 to take advantage of new opportunities and choose actions other than a_S. This results in an increment $(1 - \rho)\alpha$ relative to the no information case.

What happens to agent 2's alignment choices? In this model, it turns out that agent 2 will continue to align as if agent 1 was choosing a_S.

Lemma 1 *Agent 2 will maximise their alignment contribution by aligning to $a_1 = a_S$ regardless of agent 1's actual choice.*

Proof. Suppose that the probability that agent 1 chooses a_S is p. Suppose that agent 2 chooses to align to $a_1 = a_S$ and that agent 1 chooses $a_1 = a_S$. As agent 2 knows T_{12}, then they will choose the appropriate correctly aligned action and earn at least $m\gamma$.

Suppose now that agent 2 chooses to align to $a_1 = a_S$ and that agent 1 chooses $a_1 \neq a_S$. Because agent 2 was expecting agent 1 to choose a_S, their realised contribution then falls to 0, as they would have matched incorrectly.

Putting these together, for agent 2, if they behave as if 1 has chosen a_S, their expected alignment contribution is $pm\gamma$. The alternative is for agent 2 to choose instead their stand-alone correct action, which gives them an expected alignment contribution of 0. Thus, agent

---

[11]It would be possible to assume that the prediction is perfect with only some probability but, like our earlier simplifying assumptions, relaxing this would add complications without additional qualitative insight.

2 finds it optimal to align as if 1 is always choosing a_S. ∎

Lemma 1 demonstrates that if agent 2 chooses to maximise their alignment contribution, they will do so by continuing to align to $a_1 = a_S$, even if agent 1's choice may vary.[12]

Under perfect prediction, 1's stand-alone choice is correct if they follow the prediction. If this is done, given agent 2's inability to coordinate on that choice, what is the expected outcome?

Proposition 2 *The total payoff if agent 1 follows the prediction is:*

$$R = m(\alpha + \rho\gamma) + (1 - m)\rho\beta$$

Proof. As agent 1 follows the prediction, $E[I_{a_1=T_1}] = 1$. Agent 2 does not know 1's action. It will, however, maximise its probability of alignment (by Lemma 1) by aligning with $a_1 = a_S$. Thus, as agent 2 knows T_{12}, then $E[I_{\{a_1,a_2\}=T_{12}}] = \rho$. Given this, the expected payoff is as stated in the proposition. ∎

If agent 1 follows the prediction, this means that their action will vary with the external state. Proposition 2 shows that this makes agent 2's ability to align with that decision more difficult. This happens both for achieving the alignment contribution in a modular organisation but also for achieving the synergistic contribution, each of which are only realised with probability ρ.

To complete our analysis of the equilibrium with prediction, we need to examine whether agent 1 will choose to follow the prediction or not. If agent 1 does follow the prediction, then the organisation will adopt AI. Otherwise it will not.

Proposition 3 *In equilibrium, agent 1 will follow the AI prediction rather than set $a_1 = a_S$ if $\alpha > \gamma$.*

The proof follows from comparing the various expected returns in Propositions 1 and 2. The benefit of AI adoption is that the stand-alone performance of agent 1 is increased by $m(1 - \rho)\alpha$. However, the variability this implies for agent 1's action makes it more difficult for agent 2 to align correctly. Thus, their alignment contribution falls by $m(1 - \rho)\gamma$.

This implies that, even in a modular organisation as defined here (with $m = 1$), AI adoption creates alignment challenges. This result arises because the R depends on $mI_{\{a_1,a_2\}=T_{12}}\gamma$

---

[12]In the appendix, we demonstrate that the qualitative results do not change if agent 2 also has a direct contribution.

10

rather than just $m\gamma$. The rationale for that specification is that, even in a modular organisation, there is value to decisions in one part of the organisation aligning with others. Thus, even in the example of Amazon, if a new recommendation engine suggests products that are out of stock or more expensive to handle, those will require some coordination. If, instead, alignment is based on whether the "focal" action such as recommending the most popular product, a_S, is taken, then with probability $1 - \rho$ it will be harder (in this case, impossible) to realise γ.

It may be that, in the case of Amazon, the recommendation engine is doing a better job in matching to customers the products that Amazon already has in stock and possesses the ability to distribute as easily as any other. In this case, one might argue that γ is low or that R depends on $m\gamma$ rather than $mI_{\{a_1,a_2\}=T_{12}}\gamma$. Either of these specifications would mean that AI adoption would be more likely to occur. The point here is that the need for coordination is a constraint on AI adoption when such adoption leads to greater variability in one decision, which in turn makes it more difficult to align some other decisions in the organisation. Note, however, that the payoff generated if AI is adopted is increasing in γ. Thus, while a higher γ is a constraint on AI adoption, it is not something that an organisation that adopted AI would choose to reduce.

However, in terms of the synergistic contribution, which is more directly tied to the general understanding of modularity, Proposition 3 shows that AI adoption does not impact on the likelihood of that contribution at all. Instead, the increase in the quality of agent 1's decision raises that contribution, but it is offset by the fact that whenever agent 1 chooses something other than a_S, 2 cannot correctly align with it. Thus, there is no improvement in synergies. In other words, this type of modularity is *not* a constraint on the adoption of AI.

The question then becomes: if an organisation adopts AI, will it want to restructure itself for a higher m, as Bresnahan (2020) contends? Under AI adoption, it is preferable to have a more modular organisation if:

$$\alpha + \rho\gamma > \rho\beta \implies \frac{\alpha}{\beta - \gamma} > \rho \tag{2}$$

Note that, when $\alpha > \gamma$, the LHS of (2) is greater than the LHS of (1) as $\beta > \alpha + \gamma$. Thus, when AI adoption is an equilibrium outcome, the returns to having a more modular organisation are *higher* than when there is no prediction. Nonetheless, if β or ρ are sufficiently high, then you want to reduce modularity even if you adopt AI.

These choices for modularity are depicted in Figure 2. In the lower left, where alignment payoff γ is low and the likelihood that the focal state is correct is low, then it is optimal to have a modular organization with AI, as in Amazon's recommendation engine. The upper

Figure 2: **AI Adoption and Modularity Choice**
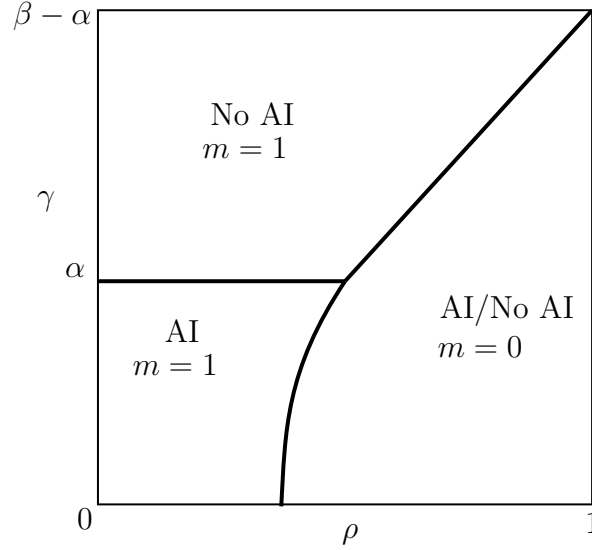
left, where alignment payoff is high and the likelihood that the focal state is correct is low, is similar to traditional retail, where recommendations and inventory choices are largely separate, AI doesn't provide decision support, and modularity is high. On the right, the focal state is likely to be correct. Both agent 1 and agent 2 benefit from choosing the focal state, getting both the alignment payoff and the stand-alone payoff, regardless of whether AI is adopted. Given that they both choose the focal state, they will choose a non-modular organization so that they capture $\beta > \alpha + \gamma$. AI provides no additional value because agent 1 will always choose the focal state. This business model is similar to the Book of the Month Club, which launched in 1926 and sent the same book to every subscriber. Since the most desired item was consistent across customers, it was possible to integrate the recommendation and shipping decisions with little risk of a mistake.

Generally, AI is only useful for improving outcomes in the bottom left quadrant, where the organisation will also choose to be modular. Modularity is best when the likelihood of the focal state is relatively low. If the focal action is relatively unlikely to be the correct action, then the AI will have an impact on the ex post decision more often. This is consistent with the intuition in Bresnahan (2020), AI is useful in modular organisations. In non-modular organizations, it offers no benefit and will not be adopted if it has a cost.

4   System Change

An interesting implication of the result that organisations that are more modular or become more modular will adopt AI is that the AI adoption choice itself impacts on a particular

decision. Thus, to the extent that AI adoption is a substitution of machine for human prediction, it is at the individual task-level. Indeed, to the extent it involves system-wide change, that change is to make the organisation operate less like a system with more weight given to operation within modules.

It is well-known that the difficulties in managing coordination across decision-makers in an organisation can be mitigated through communication (e.g., Becker and Murphy (1992), Dessein and Santos (2006)). Adoption of superior communication systems is, of course, costly. Nonetheless, the returns to adopting such systems increase as coordination becomes more difficult in their absence. The question we turn to now is whether the adoption of AI will trigger more systemic change in an organisation and drive investment in superior communication across decision-makers. If so, then arguably adopting AI will likely require architectural innovation that builds a new organisational structure from the ground up. This process takes time and is one of the reasons why adoption of general purpose technologies can be slow (David (1990)).

We focus on communication as the primary means by which system change is observed. Rather than decision-makers trying to align without a coordination mechanism, we now assume that, with a cost per message of $c > 0$, communication is possible – in particular, agent 1's prediction or their choice of action can be communicated to other decision-makers. The very fact that the prediction is generated by AI makes such communication feasible in a way that prediction undertaken by human decision-makers is not.

Suppose, therefore, that agent 1's prediction (or intended action) can be costlessly communicated to agent 2. We will assume that such communication is perfect and so the message is both received and clear.[13]

Note that such communication is of no value if there is no prediction. In this situation, as was shown earlier, agent 1 always chooses a_S and agent 2 acts as if this the case so there is no value to communicate. Indeed, having agent 1 (predictably) always choose the same action could be seen as obviating the need for communication. Thus, an organisation not adopting AI will not communicate and will not incur the cost, c.

By contrast, if agent 1 is following the AI prediction, there is such a value. To see this, suppose that agent 1 chooses to communicate an action if they know that $T_1 \neq a_S$. This happens with probability, $1 - \rho$. Given this, it is reasonable to assume that agent 2 believes that a^S will be chosen unless they receive a message to the contrary. With communication, agent 2 can coordinate on agent 1's chosen action. Thus, with communication, the expected

---

[13]Once again the model could incorporate imperfections in communication with the cost of additional notation but without any substantive qualitative insights.

payoff should AI be adopted is:

$$R = m(\alpha + \gamma) + (1 - m)\beta - (1 - \rho)c$$

Note that when $T_1 = a_S$, agent 1 need not send a message to agent 2. Thus, c is only incurred with probability $1 - \rho$.

Given this, we can prove the following:

Proposition 4 *If $m \min\{\alpha, \gamma\} + (1 - m)\beta \geq c$, agent 1 will always choose to follow the AI prediction with communication. When AI with communication is adopted, the expected payoff is always higher as m falls.*

The proof is straightforward and thus omitted. Note that AI adoption can occur even if $\alpha < \gamma$ for c is sufficiently small. Proposition 4 shows that when system change is possible, this alleviates the constraints on adopting AI.

In terms of the choice of modularity, when AI with communication is adopted, it is optimal for the organisation to be non-modular; i.e., set $m = 0$. Thus, the expected returns from the adoption of AI are $\beta - (1 - \rho)c$. From this perspective, if coordination is possible, AI adoption offers the ability to earn greater synergistic returns than is possible in its absence. This analysis demonstrates that AI adoption will be complemented by system-wide change in the form of an increase in communication.

In retail, StitchFix provides an example of such a non-modular organisation that integrates AI recommendations with inventory and shipping decisions. The AI can change the decision from the focal action often, but communication means that the full coordination benefit is realised.

5 Conclusions

This paper provides a model of AI adoption with potential system level effects. Starting with the understanding that the current excitement around AI is driven by prediction technology, we examine the decision to adopt AI in an organisation with two decisions. When the decisions are coordinated, there is an extra benefit. AI allows more precise decisions, but it makes it more difficult for the two decisions to coordinate. In this case, AI will only be adopted when the benefit of getting the first decision correct outweighs the benefit of coordination. When organisations are modular, the benefit of coordination is relatively small. When better prediction can override the default decision more often, modularity is particularly beneficial.

If, however, the organisation can change the way it operates, by enabling communication between the decision-makers, then AI adoption can have an even larger impact on outcomes in non-modular organisations. In other words, non-modular organisations benefit when the organisation changes. Put in the context of the information technology adoption literature, AI adoption has the biggest benefit in non-modular organisations, but only when there is system-wide change in the way the organisation operates.

6 Appendix: Modularity by Act

In our baseline model, D_2 creates value by aligning with D_1. Here we present a model, closer to the spirit of Van den Steen (2017), where D_2 involves a choice between pursuing a separate value contribution or aligning with D_1. Thus, we suppose that both D_1 and D_2 can create stand-alone contributions of α_1 and α_2, respectively, As before, creating α_1 requires that D_1 is "correct" in that $a_1 \in T_1$. We now assume that α_2 can be similarly generated if $a_2 \in T_2$. T_2 and T_{12} are observed by agent 2 and, in contrast to T_1, agent 2 knows T_2 with certainty. The tension for agent 2 arises because $Pr[a_2 \in T_{12}|a_2 \in T_2] = 0$. Thus, agent 2 can choose to align with D_1 or alternatively generate α_2.

When there is no prediction of T_1, we have seen that agent 1 always chooses a_S. Thus, if agent 2 chooses to align with a_S, $m\gamma + (1 - m)\rho\beta$ is generated, while if they do not do so, α_2 is generated. This means that the expected payoff becomes:

$$R = m\rho\alpha_1 + \max\{m\gamma + (1 - m)\rho\beta, \alpha_2\}$$

By contrast, if there is a perfect prediction of T_1, agent 1, if they follow the prediction, always generates a contribution of α_1 while, agent 2, if they choose to align with D_1, aligns to a_S, creating a contribution of $m\rho\gamma + (1 - m)\rho\beta$. Thus, the expected payoff if agent 1 follows the prediction is:

$$R = m\alpha_1 + \max\{m\rho\gamma + (1 - m)\rho\beta, \alpha_2\}$$

The difficulty in achieving alignment causes agent 2 to be less likely to choose to align with D_1.

Given this, note that agent 1 will choose to follow the AI prediction if:

$$\alpha_1 \geq \frac{\max\{m\gamma + (1 - m)\rho\beta, \alpha_2\} - \max\{m\rho\gamma + (1 - m)\rho\beta, \alpha_2\}}{m}$$

Thus, the extent to which agent 2 chooses to align with D_1 in the absence of prediction creates a cost to following the AI (that is, whenever, $m\gamma + (1 - m)\rho\beta > \alpha_2$). Note, however, that, as before, adopting AI creates a higher return to increasing m.

While a choice of m is a choice of *modularity by design*, agent 2's choice of whether to align or not is *modularity by act*. In other words, the adjustment to the model now creates an outcome where the agents in the organisation can choose to shield themselves from uncertainty created elsewhere (in this case, the uncertainty that arises for agent 2 when agent 1 chooses a more variable action that 2 cannot observe). If α_2 is low, this choice

does not bind, and the adjusted model is the same as our main model. But as α_2 becomes larger – reflecting the ability of other parts of the organisation to insulate themselves from variation arising from D_1 – then the models have different outcomes. In particular, if α_2 is very high, the organisation is modular by act, and so it makes sense to also be modular in design (with $m = 1$), since this enhances the realised stand-alone contribution from D_1.

Finally, suppose that a communication infrastructure can be adopted so that agent 2 is communicated T_1 at a cost of c per message. In this case, agent 2 will choose to align with D_1 if $m\gamma + (1 - m)\beta - \rho c \geq \alpha_2$. So long as c is not too high, this represents agent 2's strongest incentive to align with D_1. Moreover, this incentive is higher as m becomes lower. In this case, the expected payoff is:

$$R = m\alpha_1 + \max\{m\gamma + (1 - m)\beta - \rho c, \alpha_2\}$$

When c is low, there is always an incentive to adopt AI and if this is done, it is optimal to set $m = 0$ as long as $\alpha_2 < \beta - \rho c$.

Thus, the inclusion of modularity by act only reinforces our conclusions regarding the interaction between modularity and the adoption of AI.

References

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

Agrawal, A., Gans, J., and Goldfarb, A. (2018a). *Prediction Machines: The Simple Economics of Artificial Intelligence.* Harvard Business Press.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2018b). Prediction, judgment and complexity: A theory of decision making and artificial intelligence. *The Economics of Artificial Intelligence: An Agenda.*

Aral, S., Brynjolfsson, E., and Wu, L. (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, 58(2):913–931.

Baldwin, C. Y. and Clark, K. B. (2000). *Design rules: The power of modularity*, volume 1. MIT press.

Becker, G. S. and Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160.

Bresnahan, T. (2020). Artificial intelligence technologies and aggregate growth prospects.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117(1):339–376.

Bresnahan, T. and Greenstein, S. (1996). Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers on Economic Activity: Microeconomics*, pages 1–77.

Brynjolfsson, E. and McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Reviwew.*

Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534.

Brynjolfsson, E., Rock, D., and Mitchell, T. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers Proceedings*, 108:43–47.

Campagnolo, D. and Camuffo, A. (2010). The concept of modularity in management studies: a literature review. *International journal of management reviews*, 12(3):259–283.

Cockburn, I. M., Henderson, R., and Stern, S. (2019). The impact of artificial intelligence on innovation. *The Economics of Artificial Intelligence: An Agenda*, page 115.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *The American Economic Review*, 80(2):355–361.

Dessein, W. and Santos, T. (2006). Adaptive organizations. *Journal of Political Economy*, 114(5):956–995.

Dranove, D., Forman, C., Goldfarb, A., and Greenstein, S. (2014). The trillion dollar conundrum: Complementarities and health information technology. *American Economic Journal: Economic Policy*, 6(4):239–70.

Felten, E. W., Raj, M., and Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 108:54–57.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., and Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14):6531–6539.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280.

Gans, J. and Leigh, A. (2019). *Innovation+ Equality: How to Create a Future that is More Star Trek Than Terminator*. MIT Press.

Schilling, M. A. (2000). Toward a general modular systems theory and its application to interfirm product modularity. *Academy of management review*, 25(2):312–334.

Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482.

Van den Steen, E. (2017). A formal theory of strategy. *Management Science*, 63(8):2616–2636.

Webb, M. (2020). The impact of artificial intelligence on the labor market. *Working paper, Stanford University*.