Improving Labor Market Design to Reduce Labor Abuse in the Global Supply Chain in Southeast Asia

by

Boyu Liu

B.A. Environmental Anslysis, Mathematics (minor)
Pomona College (2016)
Submitted to the Institute for Data, Systems, and Society
and
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Technology and Policy
and
Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 2022

Author …………………………………………………………………...……....
Institute for Data, Systems, and Society
And
Department of Electrical Engineering and Computer Science
January 25, 2022

Certified by …………………………………………………………………….....
Joann de Zegher
Maurice F. Strong Career Development Professor
Assistant Professor, Operations Management, MIT
Thesis Supervisor

Certified by …………………………………………………………………….....
Constantinos Daskalakis
Professor of Electrical Engineering and Computer Science
Computer Science and Artificial Intelligence Laboratory, MIT
Thesis Reader

Accepted by …………………………………………………………...……....
Noelle Eckley Selin
Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences
Director, Technology and Policy Program

Accepted by …………………………………………………………...……....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

1

this page is intentionally left blank

Improving Labor Market Design to Reduce Labor Abuse in the Global Supply Chain in Southeast Asia

by

Boyu Liu

Abstract

Forced labor and labor abuse have become a growing concern in the global supply chain. Evidence suggest that more than 25 million people are victims of forced labor, and that much of this problem stem from the recruitment process. In collaboration with the Issara Institute (Issara), a non-profit organization based in the US and Thailand, this work aims to improve the above issue in two parts. First, we evaluate the causal relationship between inefficiency in labor recruitment and labor abuse outcomes to provide evidence-based policy suggestion. Second, we design an algorithm of joint matching and learning for the recruitment platform built by Issara, named "Golden Dreams", that aims to make it easier for workers and recruiters to find suitable matches, while using data generated by this process to estimate fair labor practices by employers. Our goal is to create employer ratings that are truth-revealing to help workers make more informed choices, and help employers meet their labor demands faster and mitigate labor risk by monitoring their labor practices on the frontline.

Leveraging 2018-2020 datasets on Myanmar-Thailand labor recruitment and worker-reported abuses, we find that an inability to efficiently alleviate labor shortages significantly worsens worker-reported abuses; an increase of one standard deviation in low-skilled labor shortages leads to a 34.5% or higher increase in worker-reported abuse in the following 2-4 weeks. Labor markets that are stressed are also simultaneously more prone to unexpected shortages and abuse. Reducing frictions in recruitment, and strengthening worker reporting mechanisms that provide near-real-time information about workplace labor abuse, are important avenues to eliminating forced labor.

As such, we collaborate with Issara on the design of a labor market to address this friction. The matching while learning part builds upon existing literature in the intersection of computer science and economics. The traditional market design literature assumes known preferences and perfect information, and the classical multi-armed bandits literature does not deal with market settings with collision of preferences and resource constraints. To develop a joint algorithm that satisfy standard axioms in a market setting, yet able to learn from historical data and leverage this learning with

uncertainty to inform future actions, require an interdisciplinary approach. We propose such a combined approach. We then discuss practical considerations when putting it into practice as well as policy and social concerns.

Thesis Supervisor: Joann de Zegher
Title: Maurice F. Strong Career Development Professor
Assistant Professor, Operations Management, MIT

Thesis Reader: Constantinos Daskalakis
Title: Professor of Electrical Engineering and Computer Science
Computer Science and Artificial Intelligence Laboratory, MIT

Acknowledgments

This thesis would not be the same without the mentors and friends I have met in the past two and half years. They have inspired and changed me both within and outside of my academic endeavor. Hence, I would like to use this opportunity to reflect on this journey, and acknowledge the numerous help, guidance, feedback, and inspiration I have received.

I have dreaded and looked forward to this moment, the moment that marks the end of my time at MIT. I know I will miss it. Its boundless opportunities for growth, vibrant community for innovation, and its spirit of humility, curiosity, and boldness, have deeply shaped me. I have learned to be not only a (proud) engineer who aspire to use his skills, technical or otherwise, to the betterment of humanity, but also a more self- and socially aware person and empathetic friend. It is where I met friends and mentors with whom I hope to remain so for my entire life. For these, I am profoundly thankful.

MIT is supported by faculty, staff, and fellow students. I feel lucky to be helped and guided by my home program and family at MIT and be shielded from the vicissitudes of life. Most notably, during COVID, I stayed in Cambridge being mostly protected and supported when COVID affected so many people in the world. I would like to thank MIT medical and all those involved in the transition to online learning.

Specific to my experience, I would like to first acknowledge the Technology and Policy Program (TPP) at the Institute of Data, Systems, and Society (IDSS) of School of Engineering. It all started with an admission letter. I appreciate that we have an intimate environment of people with diverse backgrounds, and yet all share a passion for the intersection of innovation and technologies and their impact on the biggest challenges in society. Its interdisciplinary environment and intellectual freedom allowed me to be exposed to and test out many interests of mine and pursue them with confidence and support.

- Barb, thank you for being a warm presence and adding a human touch to everything we've gone through.
- Frank, thank you for always reminding me the value of independent thinking and humor.
- Noelle, thank you for being there for me when I stressed about being an international student at this confusing time.
- Ed, thank you for being our first introduction to Boston and helping us appreciate life outside of school.

The most significant part of my life at MIT revolves around my relationship with my research advisor Joann, to whom I hold the deepest gratitude. You introduced me to the field of sustainable supply chain. You have not only shown me how good research is done, but also about leadership, passion for social impact, and turning that passion into reality. It has been a huge inspiration to see that you've learned Indonesian and started PemPem to help smallholders farmers. Thank you for believing in me and investing in me.

I would like to thank Professors Constantinos (Costis) Daskalakis and Irene Lo for your guidance to my research. Costis, your passion for research and your pursuit of wisdom will always inspire me and remind me what is important - not satrapies for sure. Irene, thank you for helping me kickstart my project, and reminding me the importance of patience and dedication in this process.

I am also grateful for all the amazing professors and mentors I've met. Thank you to Professors Munther Dahleh, Stefanie Jegelka, Tamara Broderick, Iddo Drori, Jacob Andreas, and Jim Glass for your courses on data science, machine learning, and natural language processing. They brought me joy, a deeper appreciation of these subjects, and state of "flow". Thank you, Prof. Josh Angrist, for showing me the rigor of causal inference, and Prof. Parag Pathak for introducing me to the study of market design. I've enjoyed contemplating these problems and combining them into my thesis. Thank you to Professors Danny Weitzner, Hal Abelson, and R. David Edelman for introducing me to and sparking my interest in data privacy and internet policy. Thank you, Professors Aleksander Madry and Asu Ozdaglar, for guiding us through the in depth and important discussions on data-driven decision-making and society. These discussions showed me MIT's leadership and responsibility in making sure the technologies we develop truly benefit society. It is truly amazing to reflect on all the talented and dedicated mentors I have met along my academic journey. Thank you.

Special thanks to Bill Aulet, Jinane Abounadi, Trish Cotter, Dipul Patel, Carly Chase, and Elise Strobach from the MIT Martin Trust Center and MIT Sandbox for complementing my education with grit and optimism under uncertainty. Thank you to Professor Ron Heifetz and Tim O'Brien for your courses on adaptive leadership. They have changed how I see the world around me and how I can think about where I can help. These are truly invaluable learnings for life.

To all the incredible friends I've met: I am grateful for getting to know each of you. Thanks to Nico Z, Karan, and Himani for our expeditions into the city, cooking sessions, long chats, travels, and celebrating each other on every achievement big or small. Nico, you've given me new perspectives on work and the appreciation of life. Karan, you've shown me how it's possible to be driven and available for friends at the same time. Himani, you've shown me how to build amazing

communities. Thanks to Lydia for always being spontaneous, genuine, empathetic, and positive, for vulnerability, for courage, and for our philosophical musings. You've reminded me the beauty of dreams. Thanks to Nico G for bringing me new perspectives on the importance of arts, fun, and friendship. Let's do Tour de France together one day. Thanks to Qingyang and Liang for being my hotpot buddies and talking about life. You both have amazing tenacity, and I look forward to you achieving great things. Thank you, Brandon, for being an incredible person. I always learn something from you whenever we talk, and feel inspired by how you've grown in some way every time we meet. I hope to show you around China one day and try some kimchi again. Thank you to Elwyn, Farri, Tristan, Patrick, Aaron, Joe, Ryan for being the best sports crew and buddies, and to Axelle, Erin, Nina, Sade, Cathy, Ragini, Olivia, Drake, and everyone else for our conversations and fun together. I will miss all of you.

I am grateful for having a community outside of TPP as well. Felix, I appreciate your curiosity in conversations, your self-awareness, and your discipline. You've shown me how to dream big and work hard with laser focus to achieve them. Hongyu, it inspires me to see how you balance work, hobbies, and friendship; logic, and empathy. Let's run at Hakone one day. Peiqi, Yunpeng, and Yu, thank you for our conversations about interesting ideas and life. I hope we will cross paths again soon.

Thank you to my parents for your care, guidance, and love for all these years, and for always being there when I need to talk. You've taught me patience, a positive and calm attitude under uncertainty, and always believing in something bigger than myself. I would not have been here without those. Thank you for always believing in me and being proud of me.

Finally, thank you, Feiyue, for making the last year and half the best I could ever imagine. You've made me more empathetic, more grounded, and made me believe again that I can dream and achieve it. I can never appreciate enough how you supported and encouraged me in the process. I look forward to many more years of joy, wonder, and growth together.

Table of Contents

Part I:
Low-skilled labor shortages contribute to forced labor — Evidence from worker voice technologies in Thailand and Myanmar

Chapter 1: Introduction

"*A half-dozen other captains acknowledged that forced labor is common. It is unavoidable, they argue, given the country's demand for laborers. Short-handed at the 11th hour, captains sometimes take desperate measures. 'They just snatch people,' one captain explained.*"

– Ian Urbina, New York Times. "Sea Slaves: The Human Misery That Feeds Pets and Livestock," 2015

The International Labor Organization (ILO) estimates that over 25 million people are currently victims of forced labor ([1]), an extreme form of exploitation akin to modern slavery. Forced labor is defined as work or services that a person performs under the threat of a penalty and for which the person has not offered himself or herself voluntarily.

Low-skilled migrant workers are especially prone to forced labor (see, e.g., [2][3][4]). They migrate from a different country or region to escape poverty, conflicts, or lack of job opportunities at home ([4]). As a result, their price elasticity of labor m is close to infinite; even if working conditions are poor in the destination region, they likely are (or appear) better than alternatives at home. They can also be at a higher risk of labor exploitation due to high recruitment fees ([3][4]), information asymmetry ([3]), lack of law enforcement awareness ([3][4]), immigration status and visa restrictions ([3] [4]), etc.

Concurrently, rising living standards and employment expectations in regions with growing economies lead to local shortages of low-skilled labor. As a result, documented incidents of forced labor occur in countries with relatively high per capita wealth compared to

the country of origin of the victims ([5]). For example, Tickler et al. ([6]) find that economic disparity between labor demand and labor supply countries creates fertile ground for modern slavery.

Labor organizations, supply chain professionals, multinational companies, and governments are increasingly paying more attention to solving this problem. According to a review by the International Labor Organization ([7]), efforts to end forced labor usually consist of "prevention", "protection", "remedies", and "enforcement", with prevention measures focusing on "awareness raising", "fair recruitment", "due diligence", and "addressing root causes and risk factors". For example, there are labor provisions in trade and international loan agreements, corporate social responsibility auditing programs, and liability schemes. Some countries also instituted government-to-government recruitment to uphold legal and ethical standards ([7]).

However, many of these measures may inadvertently create more friction for recruitment at the same time. Thailand is "the main destination country" of migrant workers from Myanmar - among 4.9 million non-Thai workers, "approximately 76 per cent" are Myanmar ([8]). The two governments negotiated and signed a memorandum of understanding to set up a formal recruitment channel and counter labor abuse issues occurring in this migration, which includes worker registration, permit approval, and renewal processes ([9]). But this procedure is "complicated, lengthy, and expensive" ([10]). As such, formalizing recruitment through this channel may create bigger gaps between labor supply and demand and increase the overall hiring cost and friction for employers.

Because there rarely are centralized or online job markets for low-skilled labor in destination countries, search costs are large and both employers and jobseekers rely on informal

brokers. The difficulties in efficiently recruiting low-skilled workers locally are well documented globally, across a wide range of industries and countries ([11]-[15]). Informal brokers typically burden recruitment costs onto jobseekers, and the burdening of costs and resulting debts onto migrant workers can transform the labor recruitment process into a process of human trafficking ([11]-[15]). Indeed, evidence so far indicates that much of labor exploitation in the workplace has roots in the recruitment process and the practice of deception, extortion, and debt burdening often used therein. Difficulties in meeting demand for labor might increase the use of deception in recruitment, a key means of human trafficking ([16]).

Research question

In this paper, we examine whether mismatches between demand and supply of low-skilled labor contribute to forced labor and human trafficking in the workplace. Specifically, using unique datasets, we study whether labor abuse in the workplace increases when companies face large pressures on their labor force and are unable to efficiently find relief through local hiring, i.e., they face a labor shortage.

In examining the role of labor markets in forced labor, we hope to inform pro-active, market-based interventions that can mitigate forced labor risk before workers experience harm in the workplace. This is a critical practical contribution to the anti-trafficking and ethical production communities, which have traditionally focused on identifying and remediating harms to workers after these occurred (e.g., through audits, inspections, and criminal justice). We caveat that our aim was not to develop a predictive model of forced labor, which would need to capture other, non-market drivers -- e.g., physically abusive line supervisors whose behavior is unchanged by economic conditions ([17]). Rather, we aimed to test whether labor market forces are a significant contributor.

We focus our analyses on low-skilled Burmese migrant workers in Thailand, leveraging two unique datasets from the Issara Institute, a non-profit organization advancing ethical supply chains in Asia. Economic migration from Myanmar to Thailand is the largest in the ASEAN region ([18]), and the prevalence of forced labor in Thailand is significant. Across Thailand, it was estimated that in 2018 approximately 610,000 people are living in modern slavery, an increase from 425,500 in 2016 ([19]).

The first dataset contains government-level labor recruitment data detailing weekly demands for (that is, shortages of, see chapter 3) Burmese migrant workers of companies in Thailand (January 2018 to February 2020). The second dataset contains labor abuses reported during the same timeframe by workers to a (toll-free) phone-based and instant-messaging based helpline that allows migrant workers in Thailand to request information or assistance, or report information related to working conditions (see chapter 3 for further information).

Chapter 2: Literature review

We review literature on two topics: what leads to labor exploitation, and what is being done about it. The problem of labor exploitation is a complex problem with a long history, many stakeholders and concerned parties, many challenges, as well as attempts to solve them. It is impossible to do justice to this literature in the space we have here, but we aim to present pieces we think are relevant to the core of the issue as we understand it and to our research question.

It is important to note first that the power balance between brands and supplier companies plays a huge role in labor exploitation outcomes. The burden should not be on suppliers only. A report by the Human Rights Watch extensively documented the dynamics that need to be improved ([20]). For example, brands can either demand aggressive discounts, or simply delay or refuse to pay. They may delay the mass production without adjusting the delivery date, which puts additional pressure on suppliers. The reasons for delay include modification to order, approval process, and missing details in the order. When delay happens, brands may require suppliers to pay for air shipment, adding significant economic pressure to suppliers. Things get worse when all the purchasing is done through agents instead of directly managed by the brands, as opacity and distance in the process reduces oversight. Suppliers may also engage in unauthorized subcontracting, a significant risk factor for labor exploitation situations, and all the factors mentioned before may contribute to it ([21]).

An emerging evidence base suggests that economic pressures in the workplace contribute significantly to worker abuse ([20]-[24]). Case-based research corroborated that unpredictable variability in purchase orders leads to labor abuse, as companies scramble to keep up ([22]).

When production needs elevate due to increased purchase orders, there are two complementary ways to meet them, increasing capacity through hiring and capital investments and increasing productivity of the existing workforce. The equilibrium level of the two approach depends on their relative elasticity - the easier approach will be used more. Empirical work indicates that increasing workplace productivity, e.g. through lean manufacturing, has a significant impact on forced labor outcomes ([23], [24]). On the flip side, evidence from the field also indicates that production targets can at times become unrealistic for the existing workforce, leading to forced and unpaid overtime, or unfair punishments for not reaching production targets ([17]).

Often, recruiting more workers can be hard. Cost, time, and worker availability all contribute to its difficulty ([20]). Suppliers are under immense economic pressure from the purchase orders. The modern supply chain prizes speed-to-market, a pressure passed down to supplier factories and in turn impact workers. As we shall see in our dataset, local availability of workers is often limited, which prompts the need to migrant workers. In our case of Thailand, labor shortage has recently been considered a threat to its export and economic recovery ([25], [26]).

Why don't companies increase wage to attract workers? In additional to the difficulties in attracting more workers mentioned earlier, there is another darker reason – some suppliers get away with not doing that, which increase their profit and competitiveness in the short-term. This is because migrant workers often do not have the choice to change jobs even when facing exploitative situations. Their passports can be taken away; their work visa may restrict them to a specific employer; they may also have been charged exorbitant recruitment fees that practically render them indentured servants; and they are often not protected by local laws ([3], [15], [27]).

14

The combination of these factors push the balance towards dubious ways of squeezing longer hours from existing workers instead of attracting more workers through increased wages.

Overall, increasing production capacity is often difficult, which means that companies try to meet their production goals through production targets more often. Conversely, if it were easy to hire more workers, we would see less labor exploitation due to production stress. We build on this literature and provide large-scale empirical evidence in support of the hypothesis that an inability to efficiently find relief for labor shortages significantly worsens labor abuse of low-skilled migrant workers.

What is being done? Forced labor is a serious global problem. International organizations, national governments, and corporations are increasingly paying more attention to combating it. However, efforts to address forced labor in global supply chains have focused more on identifying labor exploitation through audits and inspections and mitigating risk if found over proactive measures to prevent exploitation from occurring in the first place (see tables 1 and 2 for a summary of current efforts).

The International Labor Organization adopted Protocol of 2014 to the Forced Labour Convention of 1930, an update of the outdated 1930 treaty, upon overwhelming votes by member governments, trade unions, and employers' organizations ([28]). It reaffirms members' obligations to prevent and eradicate forced labor, provide education of rights, conduct due diligence, protect and remedy persons under forced labor, and importantly, address the root causes of forced labor risks ([29]).

A 2018 report by the International Labor Organization provides useful summary of policies, tools, and practices by all stakeholders to combat forced labor ([30]). The measures are

broken into "prevention", "protection", "remedies", and "enforcement". Prevention measures include "awareness raising", "fair recruitment", "due diligence", and "addressing root causes and risk factors". For example, there are labor provisions in trade and international loan agreements (e.g., by the International Finance Corporation of the World Bank Group), corporate social responsibility auditing programs, and liability schemes.

Table 1. **Prevention of forced labor**: key provisions of the Forced Labor Protocol and Recommendation ([30])

Thematic Area	Provisions[a]	Policy Branch	Reference
Awareness-raising	Educating and informing workers, especially those considered to be particularly vulnerable, of their labour rights to prevent abuses that can lead to forced labour.	Education, Public information	PO29, Art. 2(a)
	Educating and informing employers to prevent their becoming involved in forced or compulsory labour practices.		PO29, Art. 2(b)
Fair recruitment	Protecting persons, particularly migrant workers, from possible abusive and fraudulent practices during the recruitment and placement process.	Labour administration, Migration	PO29, Art. 2(d)
	Eliminating the charging of recruitment fees to workers.		R203, Para. 8(a)
	Requiring transparent contracts that clearly explain terms of employment and conditions of work.		R203, Para. 8(b)
	Establishing adequate and accessible complaint mechanisms.		R203, Para. 8(c)
	Imposing adequate penalties.		R203, Para. 8(d)
	Regulating or licensing recruitment services.		R203, Para. 8(e)
	Orientation and information for migrants, before departure and upon arrival.	Public information	R203, Para. 4(g)
Addressing forced labour in business operations and supply chains	Supporting due diligence by both the public and private sectors to prevent and respond to risks of forced or compulsory labour.	Industrial relations, Corporate governance	PO29, Art. 2(e)
	Providing guidance and support to employers and businesses to take effective measures to identify, prevent, mitigate, and account for how they address the risks of forced or compulsory labour in their operations or in products, services, or operations to which they may be directly linked.		R203, Para. 4(j)
Addressing root causes and risk factors — a focus on the informal economy	Respecting, promoting, and realizing fundamental principles and rights at work.	Labour relations, Labour administration, Industrial relations, Corporate governance	R203, Para. 3(a)-(c);
	Basic social security guarantees forming part of the national social protection floor, as provided for in the Social Protection Floors Recommendation, 2012 (No. 202).	Social security	R203, Para. 3(d)
	Skills training programmes for at-risk population groups to increase their employability and income-earning opportunities and capacity.	Vocational education and training	R203, Para. 3(d)

Note: (a) For the full text of the provisions, see P029 - Protocol of 2014 to the Forced Labour Convention, 1930 at http://www.ilo.ch/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:P029; and R203 - Forced Labour (Supplementary Measures) Recommendation, 2014 (No. 203) at http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:3174688

Table 2. **Protection, remedies, and enforcement**: key provisions of the Forced Labor Protocol and Recommendation ([30])

Thematic Area	Provisions[a]	Policy Branch	Reference
Protection: identification and release	Strengthening efforts to identify people in forced labour.	Justice, Labour administration	R203, para. 13(d)
	Targeted efforts to identify and release people in forced labour.		R203, para. 5(1)
	Developing indicators of forced labour for use by relevant actors.		R203, para. 13(d)
Protection: regular collection of forced labour information and statistics	Regularly collecting, analysing and making available reliable, unbiased and detailed information and statistical on the nature and extent of forced labour.	Statistics	R203, para. 2(1)
Protection: immediate assistance and long-term rehabilitation	Ensuring adequate and appropriate accommodation.	Health, Social Welfare	R203, para. 9(b)
	Provision of health care, including both medical and psychological assistance, as well as provision of special rehabilitative measures for people in forced labour, including those who have also been subjected to sexual violence.		R203, para. 9(c)
	Provision of material assistance.		R203, para. 9(d)
	Provision of social and economic assistance, including access to educational and training opportunities and access to decent work.		R203, para. 9(f)
	Protecting the safety of victims as well as of family members and witnesses, as appropriate.		R203, para. 9(a)
	Protecting the privacy and identity of victims.		R203, para. 9(e)
Protection: measures for children	Access to education for girls and boys.	Education, Social Welfare	R203, para. 10(a)
	Appointment of a guardian or other representative, where appropriate.		R203, para. 10(b)
	Ensuring a presumption of minor status, pending age verification, when the person's age is uncertain but there are reasons to believe him or her to be less than 18 years of age.		R203, para. 10(c)
	Reuniting children with their families, or, when it is in the best interests of the child, providing family-based care.		R203, para. 10(d)
Protection: measures for migrants	Provision of a reflection and recovery period in order to allow the person concerned to take an informed decision relating to protective measures and participation in legal proceedings.	Labour, Interior, Migration	R203, para. 11(a)
	Provision of temporary or permanent residence permits and access to the labour market.		R203, para. 11(b)
	Facilitation of safe and preferably voluntary repatriation.		R203, para. 11(c)

Thematic Area	Provisions(a)	Policy Branch	Reference
Protection: uncondition-ality and freedom from prosecution	Ensuring that protective measures are not conditional on the victim's willingness to cooperate in criminal or other proceedings.	Justice	R203, para. 5(2)
	Ensuring that people in forced labour are not subject to prosecution or penalties for their involvement in unlawful activities that they have been compelled to commit as a direct consequence of being subjected to forced labour.		P029, Art.4(2); R203, para. 7
Remedies, such as compensation	Ensuring that all who have endured forced labour have access to justice and other appropriate and effective remedies, such as compensation for personal and material damages.	Justice, Labour Administration	P029, Art.4(1); R203, para. 12
	Providing that those freed from forced labour can pursue compensation and damages from perpetrators, including unpaid wages and statutory contributions for social security benefits.		R203, para. 12(b)
	Ensuring access to appropriate existing compensation schemes.		R203, para. 12(c)
	Providing information and advice regarding legal rights and the services available, as well as access to legal assistance, preferably free of charge.		R203, para. 12(d)
	Providing that people who have endured forced labour, both nationals and non-nationals, can pursue appropriate administrative, civil and criminal remedies in the State where the violation occurred.		R203, para. 12(d)
Enforcement of laws and regulations	Strengthening the enforcement of national laws and regulations and other measures.	Justice, Labour Administration	R203, para. 13
	Giving to the relevant authorities, such as labour inspection services, the necessary mandate, resources, and training to allow them to effectively enforce the law and cooperate with other organizations concerned for the prevention and protection of victims of forced or compulsory labour.		R203, para. 13(a)
	Providing for the imposition of penalties, in addition to penal sanctions, such as the confiscation of profits of forced or compulsory labour and of other assets in accordance with national laws and regulations.		R203, para. 13(b)
	Ensuring that legal persons can be held liable for the violation of the prohibition to use forced or compulsory labour.		R203, para. 13(c)

Note: (a) For the full text of the provisions, see P029 - Protocol of 2014 to the Forced Labour Convention, 1930 at http://www.ilo.ch/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:P029; and R203 - Forced Labour (Supplementary Measures) Recommendation, 2014 (No. 203) at http://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:3174688.

At the national governments level, the Bureau of International Labor Affairs in the US Department of Labor publishes three reports on international child labor and forced labor.

Section 307 of the Tariff Act of 1930 (19 U.S.C. §1307) prohibits the import of any product or good that was produced wholly or in part by forced labor ([31]). As one of the biggest importers of goods, the US has significant leverage to push for changes globally. On September 15, 2021, the European Union (EU) Commission President Ursula von der Leyen also announced EU's intention to introduce a ban on products made by forced labor ([32]).

Increasingly, leading global companies are taking forced labor risks seriously. Apple, for example, claims a zero-tolerance policy toward forced labor in its supply chain ([33]). It focuses on establishing codes and standards, conducting due diligence, and participating in and help labor organizations. After a comprehensive assessment of some of its suppliers on migrant worker issues, Patagonia developed a migrant worker standard that "covered every aspect of employment, including pre-hiring interactions, labor contracts, wages and fees, retention of passports, living and working conditions, grievance procedures and repatriation." ([34]) In addition to a heavy emphasis on auditing, standard setting, and participation in industry alliances, Unilever also prides itself for going beyond compliance. It works with suppliers and partners to develop their capacity and creates grievance mechanisms to hear the voices of workers throughout their supply chain ([35]).

Countries that are sources or homes of migrant workers often instate government-to-government recruiting to combat human trafficking and forced labor. South Korea created a Employment Permit System and "signed memoranda of understanding with 15 countries of origin", which is considered "one of the most advanced models of a government-to-government recruitment system" ([30]). As mentioned earlier, the Thai and Burmese governments set up a formal recruitment channel in 2003 ([9]), but it was insufficient to satisfy the huge influx of migrant workers ([10]). There is a huge number of undocumented migrant workers in Thailand

21

(100,000 are expected to be registered in November 2021 [26]). The Thai economy was home to about 2.5 million foreign workers before the pandemic, with the highest needs from the construction and food sectors ([26]).

In the United States, a farm workers union "with approximately 10,000 members in Ohio and North Carolina", negotiated with the North Carolina Growers' Association, "the largest employer of agricultural guest workers in the United States", to obtain the right to oversee recruitment ([30]). The United Farmworkers Union started running a "recruitment enterprise within a broader ethical food initiative" ([30]), known as *CIERTO*, to "assist employers with the federal H-2A visa program" ([36]). Yet the H-2A visa restricts workers from changing employers and creates significant imbalance of power, leading to significant labor abuse issues ([27]).

Migrant workers in Southeast Asia are heavily reliant on labor intermediaries, such as recruitment agencies, or informal channels. For example, the electronics industry in Malaysia is dependent on foreign workers in production, and labor intermediaries play significant and legitimate roles in the recruitment process ([15]). They fulfill many functions throughout the employment cycle "from pre-recruitment until the workers' return to their country of origin", such as getting approval from the home country, placing job orders at the sending country, screen candidates, receiving documents and applying for visa, securing final approval from sending country, security check and transportation, managing work permit and renewals, and facilitating contract termination and return to countries of origin. In the Tuna industry in the Philippines, recruitment is similarly done through either personal relationships or labor intermediaries ([37]).

Due to perils in recruitment of migrant workers, there start to be projects tacking specific problems in this process. The Issara Institute ([17]), the Unseen Project ([38]), and the Polaris Project ([39]) respectively operates worker voice hotlines in Southeast Asia, the UK, and the US. The ILO, joint with the EU, is developing a framework to mitigate problems in migrant labor recruitment that focuses on improved the institutional capacities and "accessibility to accurate information, knowledge, tools on fair recruitment" ([40]).

Chapter 3: Data and Model

Data sources

Issara Institute is a non-profit organization founded in 2014 focused on worker voice, technology and partnerships to eliminate labor abuses, including forced labor and human trafficking, from global supply chains. We used two primary datasets compiled by the Issara Institute that cover the period January 2018-February 2020.

The first dataset on labor shortages arises from official government data on demand from Thai employers for Burmese workers. The second contains data on worker-reported labor abuse, arising from worker voice technologies operated by Issara Institute. These worker voice channels, comprised of a toll-free helpline and a range of other multilingual channels leveraging the preferred technologies and interfaces of Southeast Asian migrant workers (a Yelp-like application and social media and messaging applications such as Facebook Messenger, Line, and Viber), allow migrant workers in Thailand to request information or assistance, or report information often related to working conditions, in their own words and at any time. The mechanism received over 85,000 calls and messages from workers and trafficking survivors in 2019, which drove the remediation of over 45,000 cases of labor exploitation - 8,448 of them forced labor (*29*). For perspective, this is around eight times the volume of calls and messages received by the United Kingdom national anti-trafficking hotline (*30*) and 20 times the traffic received from workers and survivors by the United States national anti-trafficking hotline (*31*) that same year. A product of this worker reporting-centered labor monitoring mechanism is a rich dataset elucidating ongoing working conditions across a wide range of industries, localities, and workplaces - many of them within the supply chains of global brands and retailers - and a

subset of this data is used for this study (see section on Worker-Reported Labor Abuse below for further information).

We also downloaded data on the Thai Baht - Chinese Yuan dollar exchange rate from the Bank of Thailand. We combined the labor shortage and worker voice datasets based on province and week. We removed provinces that either have no labor shortage or no worker voice data in any of the three years. This yielded a panel dataset with 6,554 rows (58 provinces across 113

weeks, from week 2 of 2018 to week 9 of 2020). Figure 1 shows the 58 provinces included in our

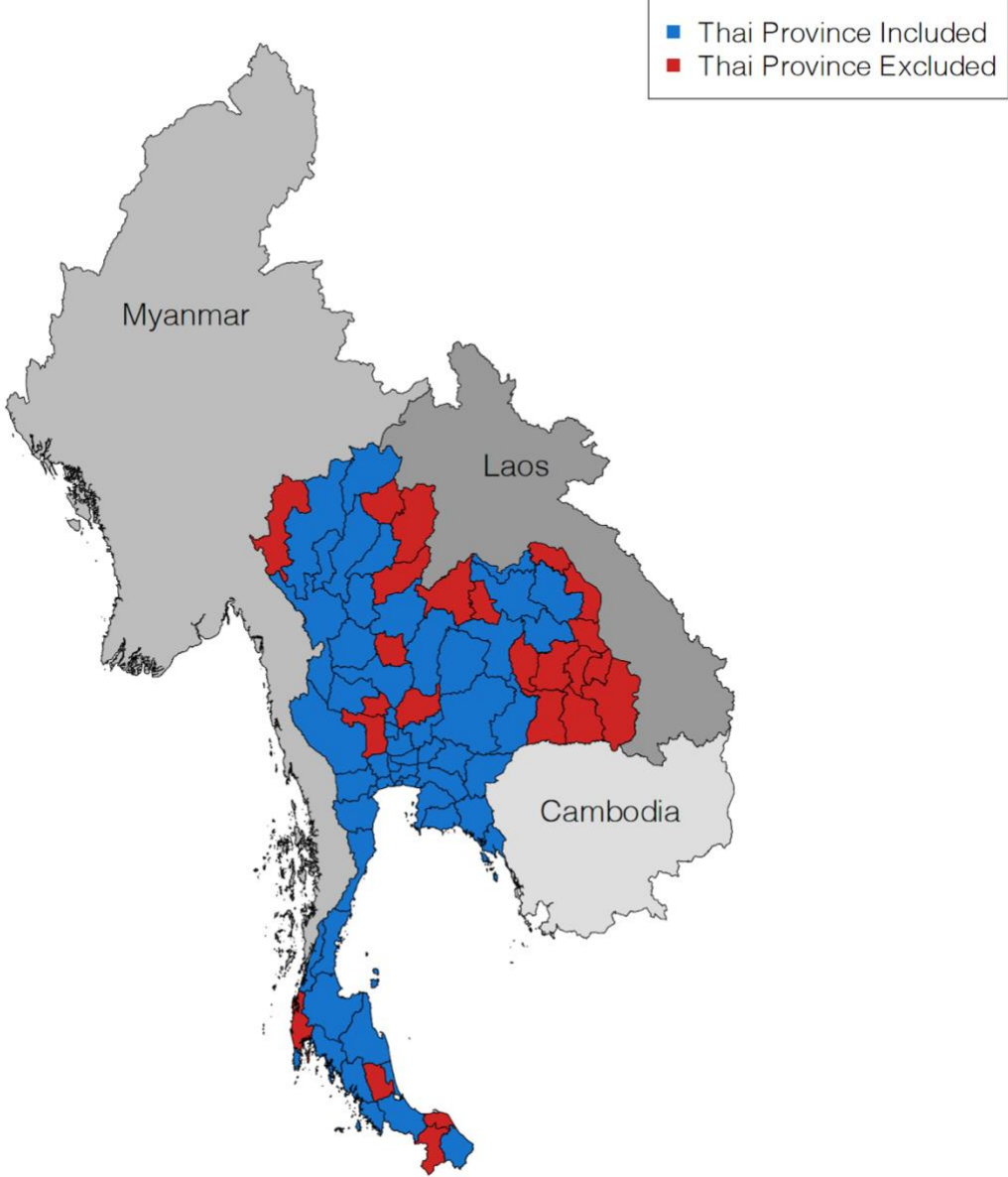analyses (blue) and 14 provinces excluded (red).

Fig. 1. Thai provinces included (blue) and excluded (red) from our analyses.

Labor shortages

We measure labor shortages through formal demands placed by Thai companies for new

Burmese migrant workers not yet in Thailand, *i.e.* Burmese prospective migrants that need to be

recruited internationally with a visa and work permit.

These formal demands capture demand for workers that is not satisfied through local

worker supply, *i.e.* through either migrant workers already in Thailand or Thai workers. Formal,

international demand for workers is a more complicated, bureaucratic, and slower process to

recruit workers than recruiting workers -- Thai or migrant -- through local informal channels and

networks (see a high-level schematic representation of the MOU process in Fig. 2). As a result,

an employer will place a formal demand to recruit non-locally only when supply is not available

locally. This formal demand therefore reflects the portion of labor demand for which there is a

local labor shortage.

Fig. 2. High-level schematic representation of Thailand-Myanmar MOU process for labor recruitment. "DOL" is the Thai Department of Labor; "MOLIP" is the Myanmar Ministry of Labor, Immigration and Population; "RA" is the Recruitment Agent; "DOE" is the Thai Department of Employment of the Ministry of Labor.

The formal process requires Thai companies to follow the steps outlined in the

Memorandum Of Understanding (MOU) between Thailand and Myanmar (Fig. 2). Once the

request for workers is approved by the Myanmar government, the number of workers demanded

is recorded. Issara Institute, in collaboration with the Myanmar government, digitized this data for January 2018 to February 2020, after which the MOU labor recruitment process was closed due to the COVID-19 pandemic.

The dataset contains formal demands for migrant labor placed by 8,675 unique companies that are located across 72 of the 77 provinces in Thailand.

Moreover, we are interested in studying the impact of unexpected shocks in labor shortages. Suppose the purchasing orders are growing steadily, we can imagine employers anticipating and adapting for the growth through various methods, such as keeping an extra capacity of workers, machines, and raw materials, and making efficiency improvements. If an employer has to hire the same number of migrant workers every month, it will become better at it. If demand deviates from the past unexpectedly, however, it would be much harder to prepare and causes labor *stress*. From figure 4, we see that demand indeed has huge swings between weeks. Also, there is substantial heterogeneity across provinces, especially in terms of size and volatility of demand (see the standard deviation in Fig 4, and Fig. 5), so we standardize the data accordingly. We do not see a clear trend in the overall labor demand (Fig. 3). Thus, we analyze standardized shocks in labor shortage; let $S_{i,t}$ denote the observed labor shortage in province i,

week t, and let $D_{i,t}$ denote the (standardized) shock in labor shortage for province i, week t.

Then:

$$D_{i,t} = \frac{S_{i,t} - \frac{1}{4}\sum_{k=1}^{4} S_{i,t-k}}{\sigma_{S_i}} \text{ for } t > 4 \qquad [1]$$

Where σ_{S_i} is the standard deviation of labor shortage in province i. Here, the shock in labor

shortage is thus measured relative to the 4-week moving average of labor shortage; we also study

the robustness of our results using the 8-week moving average of labor shortage (see Table S2 in SI.).
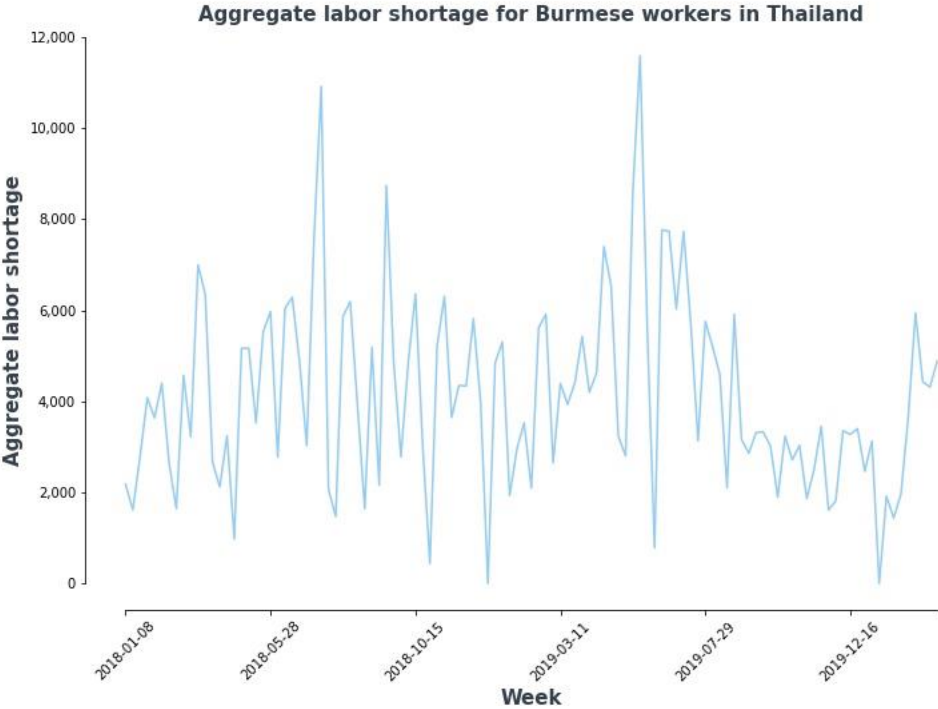

Fig. 3. Aggregate low-skilled labor shortages (in number of workers demanded internationally) over time across all provinces and companies, from January 2018 - February 2020. Source: Issara Institute.
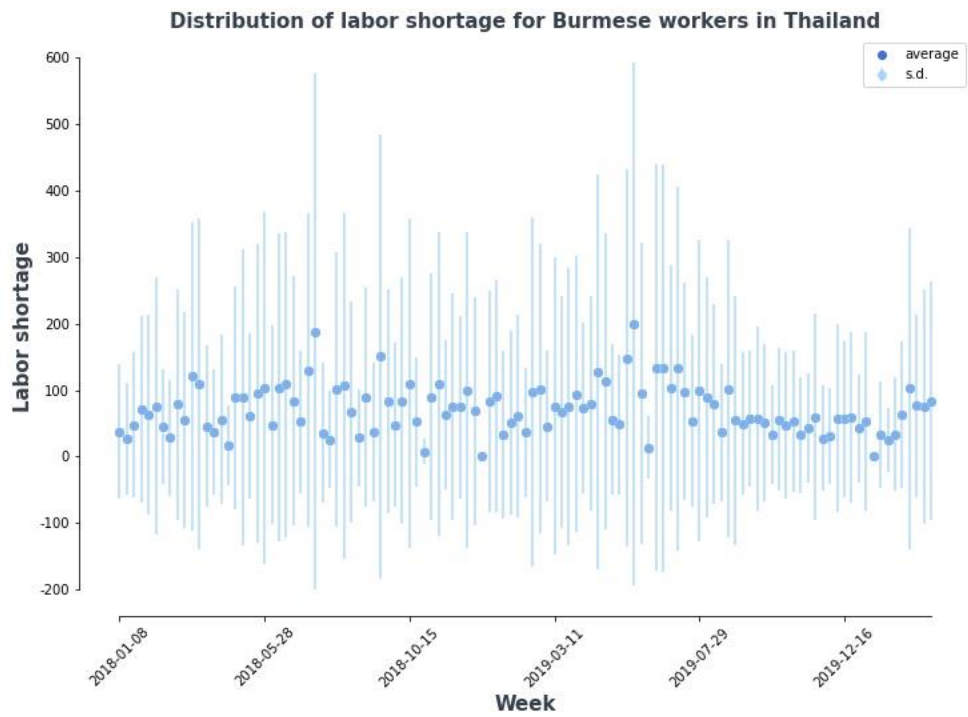
Fig. 4. Mean and standard deviation of low-skilled labor shortage (in number of workers demanded internationally) across provinces, from January 2018 - February 2020. Source: Issara Institute.
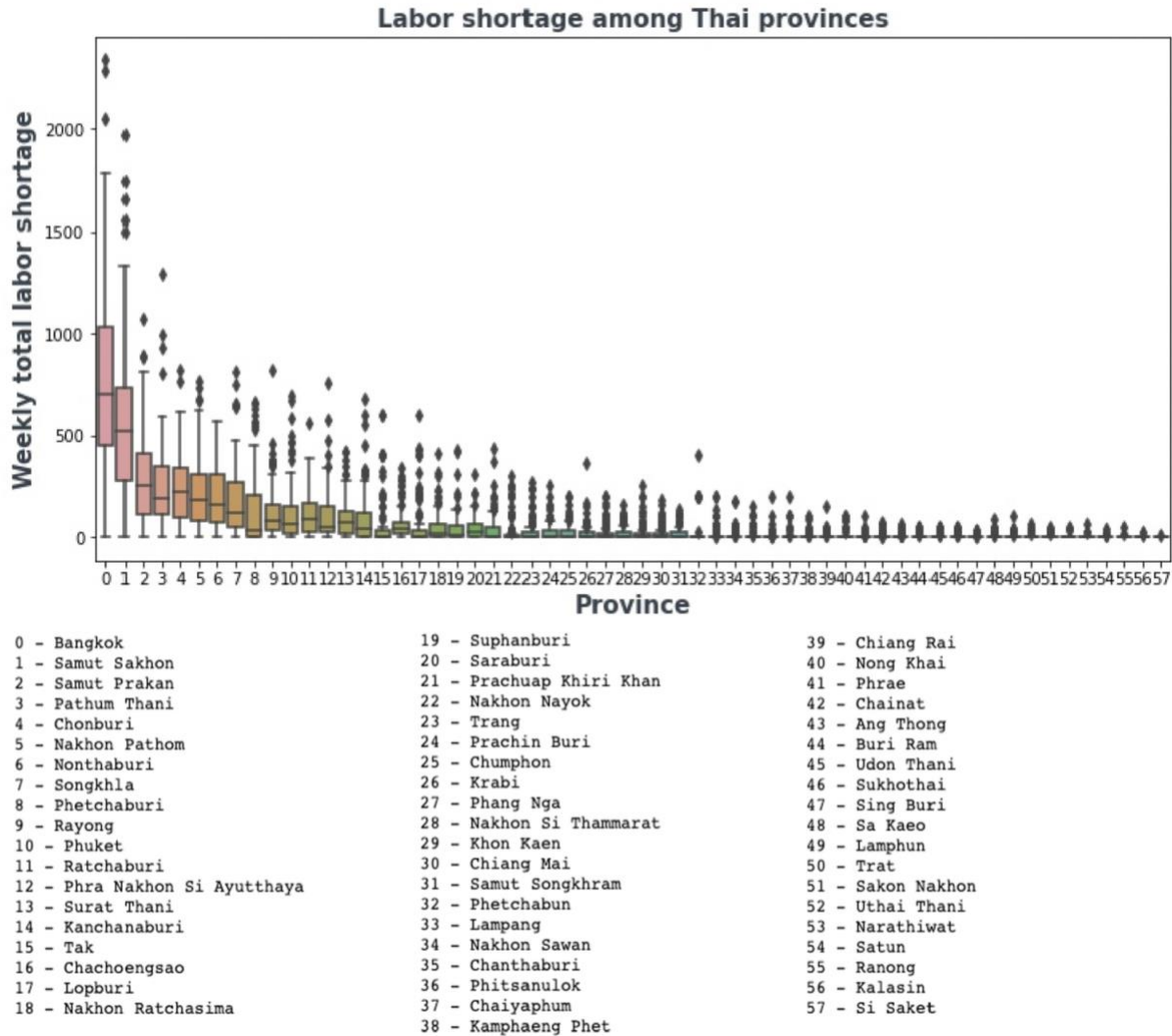
Labor shortage among Thai provinces

0 – Bangkok	19 – Suphanburi
1 – Samut Sakhon	20 – Saraburi
2 – Samut Prakan	21 – Prachuap Khiri Khan
3 – Pathum Thani	22 – Nakhon Nayok
4 – Chonburi	23 – Trang
5 – Nakhon Pathom	24 – Prachin Buri
6 – Nonthaburi	25 – Chumphon
7 – Songkhla	26 – Krabi
8 – Phetchaburi	27 – Phang Nga
9 – Rayong	28 – Nakhon Si Thammarat
10 – Phuket	29 – Khon Kaen
11 – Ratchaburi	30 – Chiang Mai
12 – Phra Nakhon Si Ayutthaya	31 – Samut Songkhram
13 – Surat Thani	32 – Phetchabun
14 – Kanchanaburi	33 – Lampang
15 – Tak	34 – Nakhon Sawan
16 – Chachoengsao	35 – Chanthaburi
17 – Lopburi	36 – Phitsanulok
18 – Nakhon Ratchasima	37 – Chaiyaphum
	38 – Kamphaeng Phet

39 – Chiang Rai
40 – Nong Khai
41 – Phrae
42 – Chainat
43 – Ang Thong
44 – Buri Ram
45 – Udon Thani
46 – Sukhothai
47 – Sing Buri
48 – Sa Kaeo
49 – Lamphun
50 – Trat
51 – Sakon Nakhon
52 – Uthai Thani
53 – Narathiwat
54 – Satun
55 – Ranong
56 – Kalasin
57 – Si Saket

Fig. 5. Boxplot distribution of weekly labor shortage across Thai provinces from January 2018 - February 2020. The boxplot shows the median and interquartile range; whiskers show observations falling within 1.5 times the interquartile range. Source: Issara Institute.

Worker-Reported Labor Abuse

We measure worker-reported labor abuse by using worker voice data from the Issara

Institute, which runs a range of worker voice channels, including a multi-lingual hotline,

Facebook, and smartphone-based chat applications such as Line, Viber, and WhatsApp. These

worker voice channels from Issara are the ones most prevalently used by migrant workers in

Thailand and receive more traffic than that of the United States and United Kingdom national

anti-trafficking hotlines combined. The worker voice data used in this study includes records of workers calling the Issara multi-lingual migrant worker hotline to report abusive conditions, to seek help/support, and/or to request information about their rights. In addition to reports on the Issara hotline, we also consider the reporting made through Issara's private messaging in Facebook and chat apps. 79% of our records are from the Issara hotline, 20% are from Issara's Facebook messages and 1% come from other sources.

Each worker voice log includes, among others, information about the time of the call/report, the location and province of the caller, the nationality of the caller, and the type and severity of situation that the hotline operator determines the caller to be in (for example "Debt bondage", "Issue with wages: unethical but not illegal", "Issue with deductions: illegal and/or excessive"). While 77.6% of the calls logged contain information about the province, only 33.8% also contain information about a company and/or industry. We therefore aggregated the data at the province level and discarded hotline calls when no province information was available. This led to a dataset of 9,182 hotline calls, of which 2,178 indicated a clear worker-reported labor abuse (see Table 3), which is 23.7% of the total. It is noteworthy that the audits to which these workplaces are routinely subjected failed to identify the labor abuses that worker-inclusive labor monitoring identified.

Table 3. Overview and examples of worker voice categorization.

"Serious labor abuse" (n = 3,376)
● Treated poorly and/or threatened by RA or broker: serious violation ● Document retention ● Debt bondage ● Inaccurate, misleading, or poor information about job: serious, deceptive misinformation ● Proper, legal documents not provided to worker ● Workers placed in workplace and jobs different from what they were recruited for

● Issue with deductions: illegal and/or excessive
● Very late or missing payments
● Unfair dismissal
● Child labor issue

"Labor abuse, less serious" (n = 797)

- Contract provided was not clear (insufficient detail about wages, benefits, job description, etc.)
- Contract provided was not in native language
- Issues with regular shift hours or other shift issues: dispute without violation of law or employment contract
- Issue with holidays or other benefits: dispute without violation of law or employment contract
- Social Security issue (access to health care): minor issue
- Weak company policies and/or poor communication of company policies to workers
- Issues with workforce and workloads: no violation of law or employment contract
- Unresponsive grievance mechanism / no action being taken

"Informational call; no KPI" (counted as no serious labor abuse) (n = 6,294)

- Caller would like to know if the social security fund (SSF) expired or not after stop contributing for three months
- Caller would like to know if the border is closing or not in 1st May and the government plan for migrant returning
- Caller asking about passport process, safe migration, worker rights, Issara Intro, GD and including COVID-19
- Client informed back she is OK now about passport processes for working with the recruitment agency
- Caller would like to know types of cancellation letter

Finally, we aggregated all observations at the weekly level to align the dataset's temporal

granularity with the temporal granularity of the labor shortage dataset. Figure 6 shows the

aggregate number of weekly worker voice reports from Jan 2018 to Feb 2020, and Figure 7

shows the distribution of percentage of worker-reported labor abuse across provinces over time.
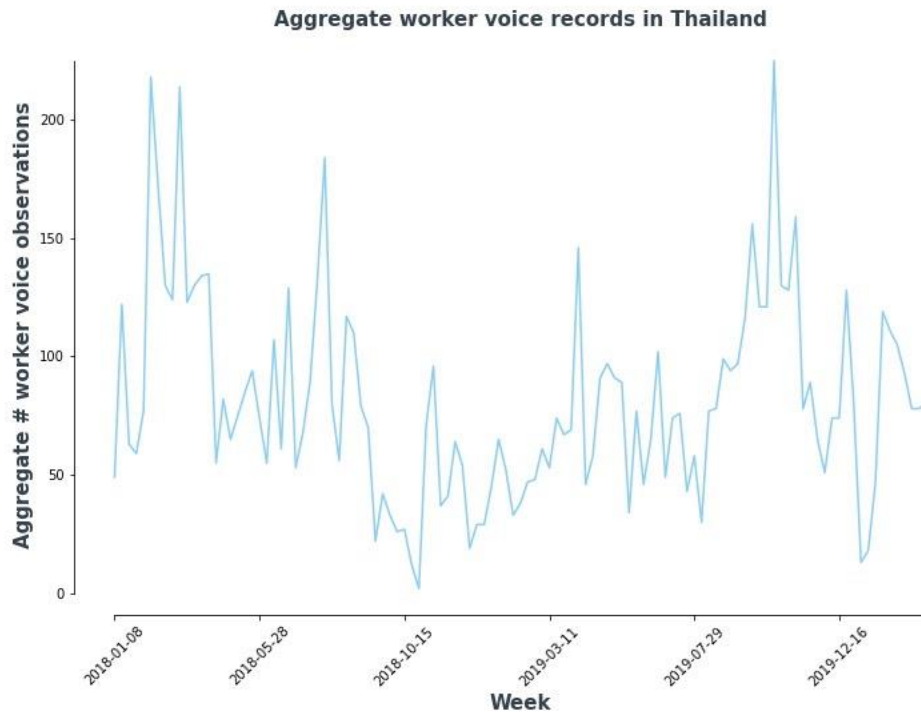

Aggregate worker voice records in Thailand

Fig. 6. Aggregate number of weekly worker voice reports (hotline only), Jan. 2018 to Feb. 2020. Source: Issara Institute.
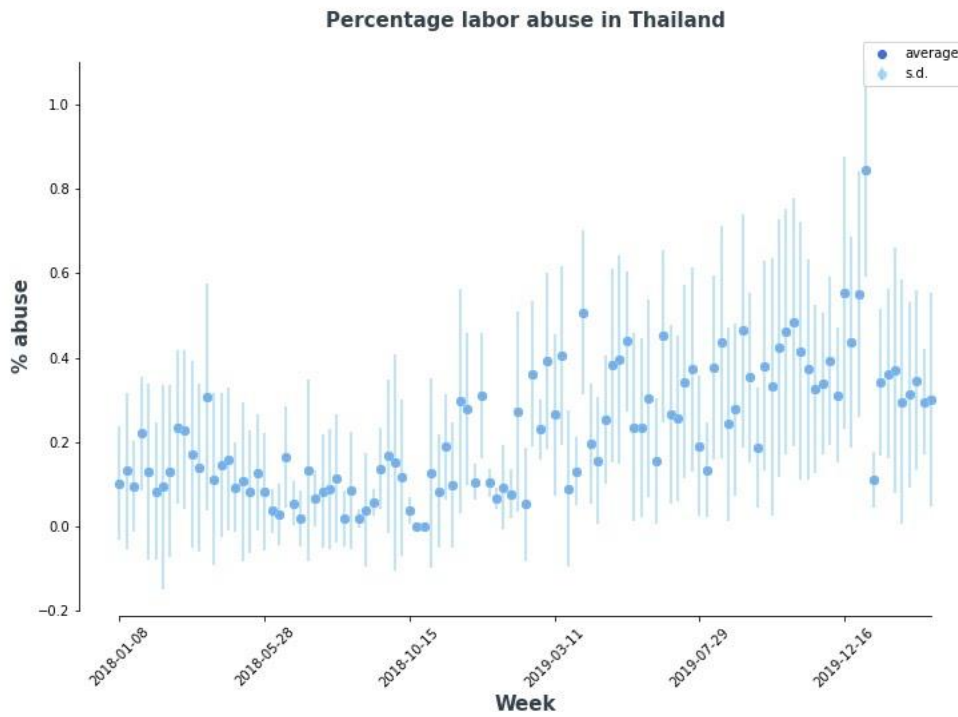

Fig. 7. Distribution of percentage of hotline calls reporting worker abuse across provinces in Thailand from Jan. 2018 to Feb. 2020. Source: Issara Institute.

The overall distribution of Burmese migrant workers and worker voice data across Thai provinces is uneven (see Figures 6-7), and the likelihood that a worker reports an abuse to Issara in a particular province depends on the relative penetration of Issara in that province. Hence, using the absolute number of calls in a province would be a biased measure of worker-reported labor abuse. To correct for the heterogeneous penetration and migrant population size, we compute the *percentage* of calls that report clear exploitative situations in each province and each week; this captures a standardized measure of abuse (in % terms) in the migrant population of a province. To do so, we first classified the calls into workers being in a "serious abusive

condition" (S), versus workers not clearly being in a serious abusive situation or calling about issues unrelated to workforce labor abuse (NS) (see Table 3).

The percentage of labor abuse is then computed as shown in Eq. 2 for province i, week t.

$$P_{i,t} = \frac{\sum_j (Call, S)_{i,t,j}}{\sum_j (Call, NS)_{i,t,j} + \sum_j (Call, S)_{i,t,j}} \qquad [2]$$
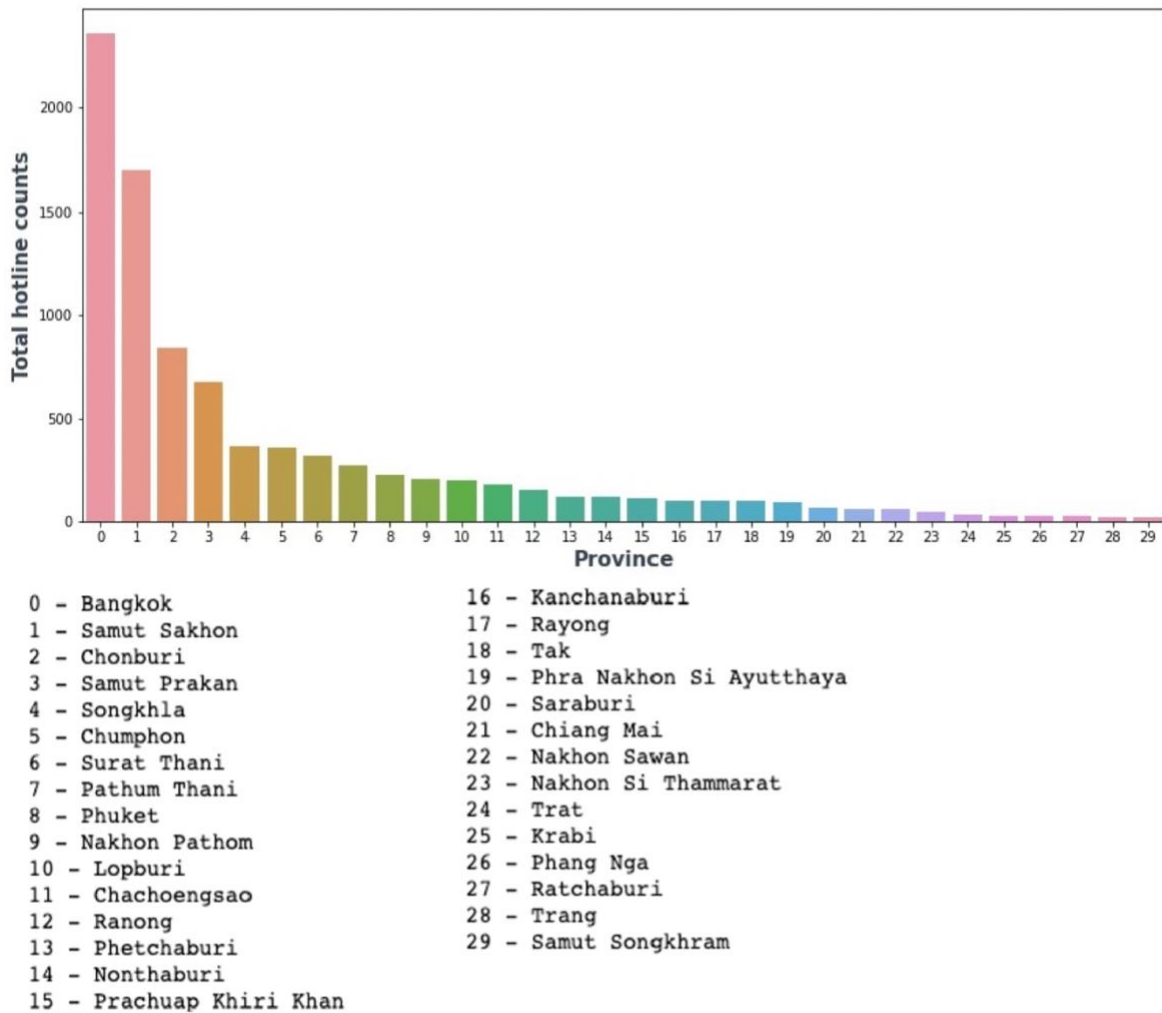

0 – Bangkok	16 – Kanchanaburi
1 – Samut Sakhon	17 – Rayong
2 – Chonburi	18 – Tak
3 – Samut Prakan	19 – Phra Nakhon Si Ayutthaya
4 – Songkhla	20 – Saraburi
5 – Chumphon	21 – Chiang Mai
6 – Surat Thani	22 – Nakhon Sawan
7 – Pathum Thani	23 – Nakhon Si Thammarat
8 – Phuket	24 – Trat
9 – Nakhon Pathom	25 – Krabi
10 – Lopburi	26 – Phang Nga
11 – Chachoengsao	27 – Ratchaburi
12 – Ranong	28 – Trang
13 – Phetchaburi	29 – Samut Songkhram
14 – Nonthaburi	
15 – Prachuap Khiri Khan	

Fig. 8. Number of hotline calls across provinces in Thailand from Jan. 2018 to Feb. 2020, among provinces with more than 20 hotline calls. Source: Issara Institute.
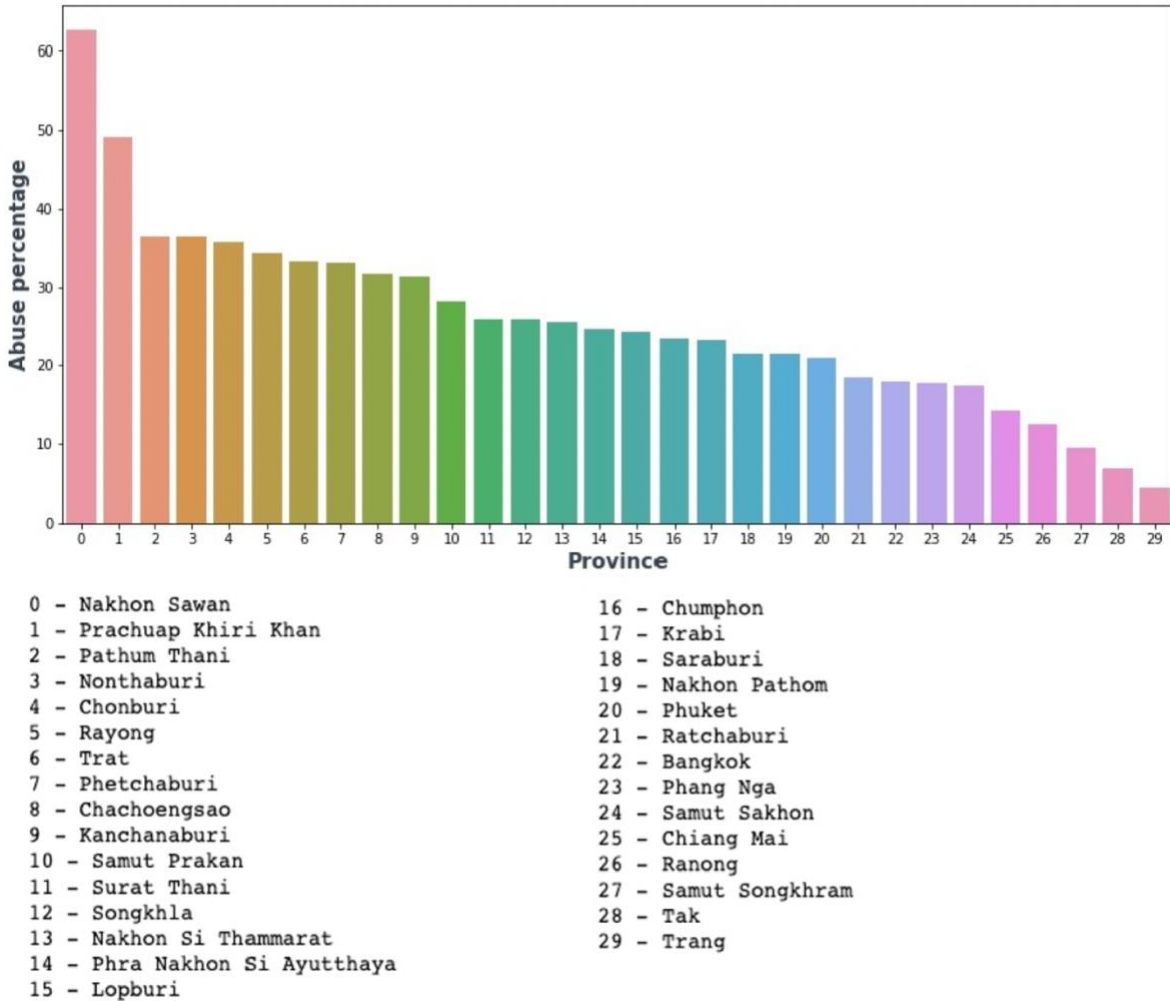
Fig. 9. Percentage of hotline calls reporting labor abuse across provinces in Thailand from Jan. 2018 to Feb. 2020, among provinces with more than 20 hotline calls. Source: Issara Institute.

Exchange rate

The exchange rate data is retrieved from the Bank of Thailand historical foreign exchange rate website (*28*). We downloaded the 2018-2020 daily "transfer" rates for the Chinese Yuan-Thai Baht exchange rate and US Dollar-Thai Baht exchange rate. China is the main trading partner of Thailand, both as an importer (11.95% of Thailand's export) and as an exporter (20.05% of Thailand's import) (*24*), so our base model uses the CYN-THB exchange rate as an instrumental variable. In robustness analyses, we also used the USD-THB exchange rate because

the US is the second largest market of Thailand, accounting for 11.14% and 6.10% of Thailand's exports and imports respectively (*24*).

We converted the daily data to a weekly time series by taking the average of daily exchange rates within each week. The Thai Baht has significantly appreciated over the past three years. Our analyses use the first differences to avoid spurious results.
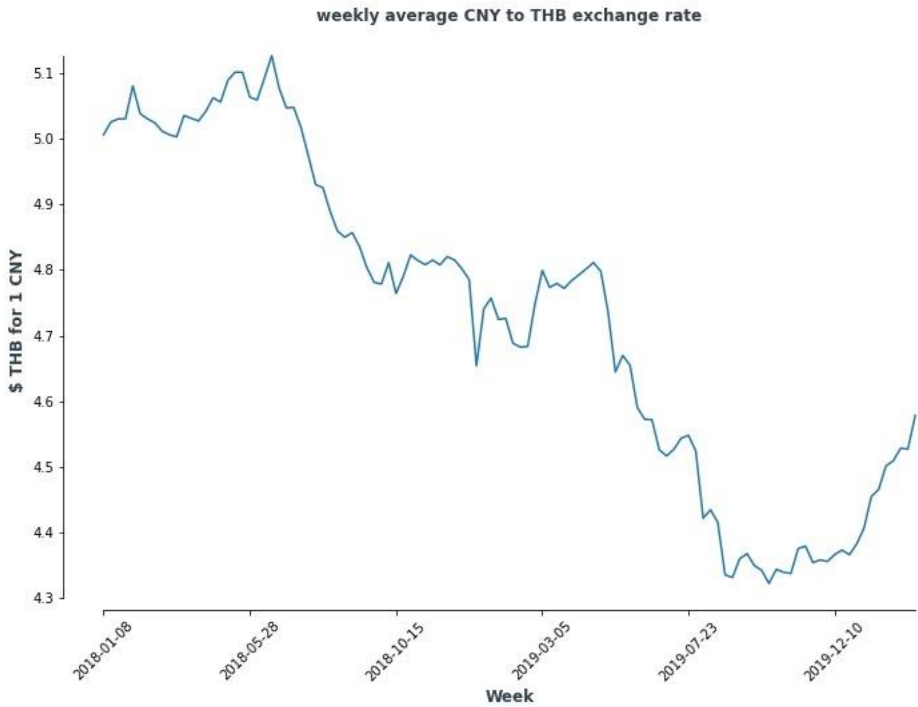

Fig. 10. Chinese Yuan-Thai Baht exchange rate, averaged at weekly level, from Jan. 2018 to Feb. 2020. Source: Bank of Thailand.
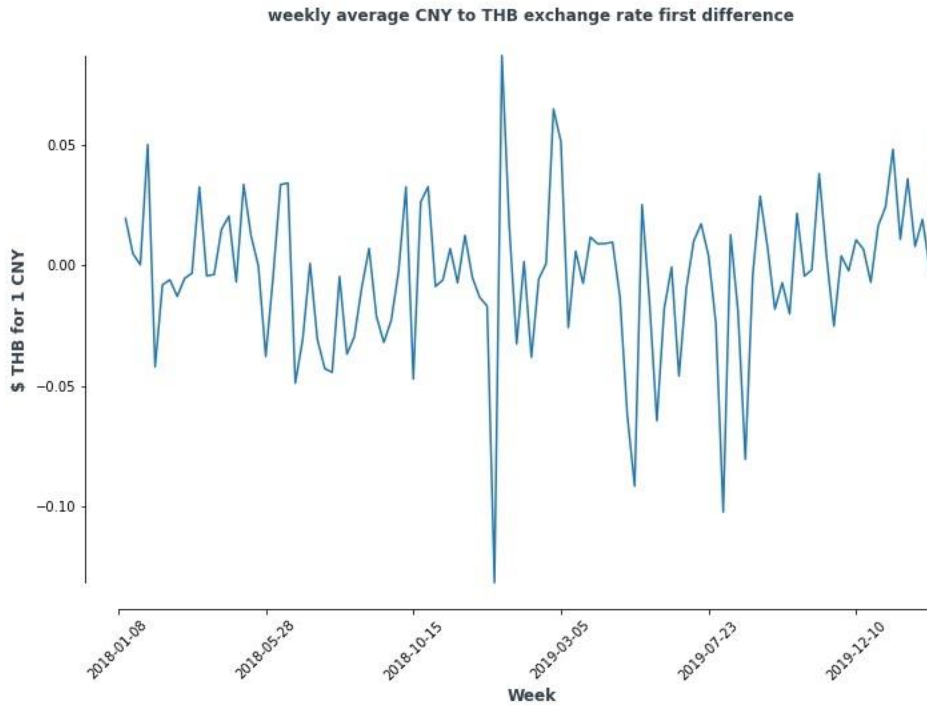
Fig. 11. First difference of Chinese Yuan-Thai Baht exchange rate, averaged at weekly level, Jan. 2018 to Feb. 2020. Source: Bank of Thailand.

Econometric Model

To estimate the impact of shocks in low-skilled labor shortages on labor abuse, we first apply the simple regression in Eq. 3 to the observational data:

$$P_{i,t} = \beta_0 D_{i,t-x} + \beta_j \overline{L} + \beta_m \overline{T} + \varepsilon_{i,t} \quad [3]$$

where $P_{i,t}$ is the percentage reported abuse (defined in Eq. 2) in province i, week t; $D_{i,t}$ is the shock in low-skilled labor shortage (defined in Eq. 1) in province i, week t; $\overline{L}$ is a vector of 58 fixed effect dummy variables for locations (provinces); $\overline{T}$ is a vector of fixed effect dummy variables for time periods (weeks, months, or quarters); and $\varepsilon_{i,j}$ is the error term. Thus, β_0 is the

effect of shocks in low-skilled labor shortage on percentage worker-reported labor abuse, our parameter of interest.

A key concern with using observational data for causal inference is that labor shortages might be endogenous. For example, one could imagine that labor abuse in a workplace drives labor turnover which in turn drives labor shortage. We remedy this by using an Instrumental Variable (IV) approach to estimate the causal effect of shocks in labor shortages on labor abuse.

The IV we use is shocks in the foreign exchange rate, specifically with China -- Thailand's main trade partner by both import and export (*24*). Shocks in the foreign exchange rate are predicted to affect shocks in labor shortages by changing the underlying economic conditions firms face, by making export of goods cheaper (more expensive) for foreign buyers, thus inducing more (less) demand for goods. A higher demand for goods increases production demand and, therefore, demand for labor at the factories. We acknowledge that there are other factors such as lead time and payment terms that also play a role in this dynamic. This is illustrated by the first three horizontal arrows on the top left in Figure 12. Field interviews confirmed that companies face stronger demand for labor when they face larger commercial pressures to produce by customers. Shocks in the foreign exchange rate is thus a promising IV, and this is corroborated by IV tests we conducted. Since the exchange rate is likely to have a lagged effect on economic conditions faced by firms, we used a lagged exchange rate as IV.

We also allow for a lag *x* between labor shortage and labor abuse, to capture the lag between high demands on the workforce and reported labor abuse and eliminate potential reverse

causality issues; while labor abuse might drive turnover in the same period or future periods, such turnover does not induce a higher labor shortage in past periods (Fig. 12, upper right cross).
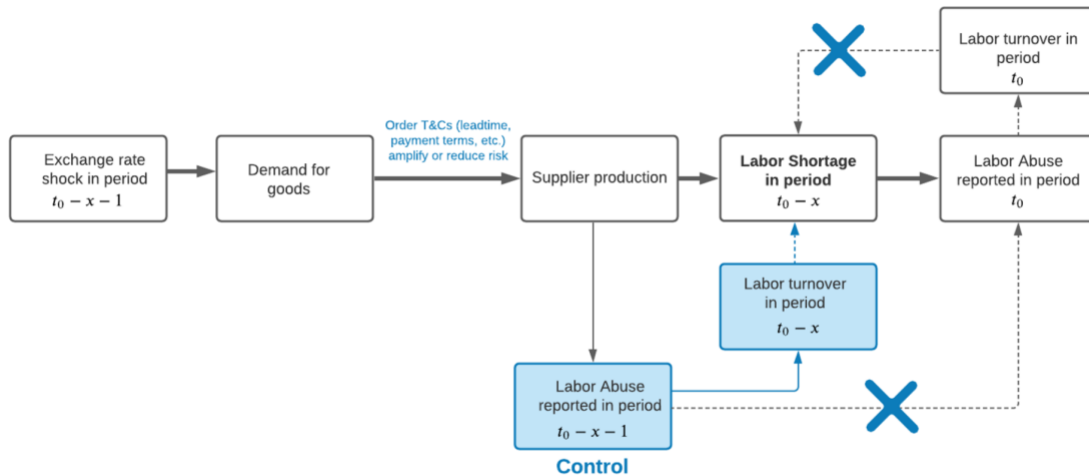

Fig. 12. Causal graph for data-generating process (x≥0)

Another concern that one might have is that increased demand for goods might *immediately* increase labor abuse for the existing workforce (*e.g.* pressure and overtime), which affects labor turnover, and thereby affects labor shortages in the next period (see the pathway in the blue boxes at the bottom of Figure 12). Such worker-reported abuse may also persist over time for various reasons (such as habits, delays in demand or delays in reporting abuse), which means that exchange rate shocks from period $t_0 - x - 1$ can also affect abuse in period t_0 (lower right arrow in dotted line in Figure 12). We address this problem by controlling for lagged abuse from period $t_0 - x - 1$ in our 2-SLS IV regression. After controlling for lagged abuse, past

exchange rate shocks are no longer correlated with abuse in period t_0 (Figure 12, lower right cross), and our IV becomes valid.

The two stages of the 2-SLS are thus:

$$D_{i,t-x} = \beta_E E_{t-x-1} + \beta_j{}^1\overline{L} + \beta_m{}^1\overline{T} + P_{i,t-x-1} + \eta_{i,t} \qquad [4]$$

$$P_{i,t} = \beta'{}_0 \widehat{D_{i,t-x}} + \beta_j{}^2\overline{L} + \beta_m{}^2\overline{T} + P_{i,t-x-1} + \varepsilon_{i,t} \qquad [5]$$

We have a first stage equation for shocks in labor shortage, where E_{t-x-1} is the exchange rate shock in week $t - x - 1$, and $\eta_{t,j}$ is the first stage error term. $\widehat{D_{i,t}}$ is the fitted value from the first stage equation. The definitions of $P_{i,t}$, $D_{i,t}$, $\overline{L}$, $\overline{T}$, and $\varepsilon_{i,j}$ are the same as the OLS set-up in Eq. 3. We use cluster-robust standard errors, clustered at the province level, for all regressions.

Chapter 4: Results

Main results

Table 4 reports our main result, under ordinary least squares (OLS) estimates and under instrumental variable (IV) estimates using Two-Stage Least Squares (2-SLS). The reported coefficient captures the effect of a shock in labor shortage on worker-reported labor abuse four weeks later (*i.e.* a lag of four weeks, or x=4 in Eq. (3)-(5)). Table 7 shows results for lower or higher lags; Tables 4, 5, and 7 provide further results for, respectively, different province inclusion criteria, different shock calculations, and the THB-USD exchange rate as IV rather than the THB-CNY exchange rate.

	OLS	2-SLS IV		
	Coeff.	Coeff.	Weak Inst.	Wu-Hausman
With province FE	<0.001 (0.002)	0.187*** (0.064)	<0.001	<0.001
With province & quarter FE	-0.002 (0.002)	0.150** (0.060)	<0.001	0.002
With province & month FE	-0.001 (0.001)	0.113** (0.055)	0.001	0.009

Outcome: % abuse in a week, on a 0 to 1 scale
Cluster Robust standard errors (province): Yes
IV: THB- CNY exchange rate, first-difference, lagged by one week

Table 4. OLS and 2-SLS IV regression results for weekly % labor abuse as a function of labor shortage shocks four weeks earlier (x= 4), for different regression specifications, and *p*-values for Weak Instruments and Wu-Hausman tests. The OLS regression specification is given in Eq. (3); the 2-SLS regression specification is given in Eq. (4)-Eq. (5). *, **, and *** denote (two-tailed) significance at the 10%, 5% and 1% level, respectively, for regression coefficient estimates. Standard errors are given between round braces.

Lag in labor shortage shock (x)	IV tests rejected?	Coeff.	Std. Error
0	False	0.005	0.034
1	False	0.043	0.033
2	True	0.082*	0.042
3	True	0.088*	0.045
4	True	0.187***	0.064

Outcome: % abuse in a week, on a 0 to 1 scale
Province fixed effects included: Yes
Time fixed effects included: No
Cluster Robust standard errors (province): Yes
IV: THB- CNY exchange rate, first-difference, lagged by one week
IV tests: weak instruments test and Wu-Hausman test, alpha=0.05

Table 5. 2-SLS IV regression results for weekly % labor abuse as a function of labor shortage shocks in preceding weeks. The 2-SLS regression specification is given in Eq. (4)-Eq. (5). *, **, and *** denote (two-tailed) significance at the 10%, 5% and 1% level, respectively, for regression coefficient estimates.

We find a consistent effect: for every lag between a shock in labor shortage and worker-reported labor abuse from two to four weeks (*i.e.* x=2,3,4 in Eq. (3)-(5)), the exchange rate in the previous week is a valid instrument, and for all lags with valid instruments, the effect of interest is significant and increasing with each lag, up to x=4. The coefficient of the exchange rate shock in the first-stage regression (β_E in Eq. 4) is positive and significant (*p<0.001*), indicating that a positive exchange rate shock (devaluation of the Thai Baht) leads to a positive shock in labor shortage -- as one would expect from economic theory.

We note that we found inconclusive results when using small or no lags in the reported labor shortage shock relative to worker-reported labor abuse (*i.e.* x=0 or x=1, see Table 5). In the 2-SLS IV regression, this lag can be thought of as being composed of two types of lags: first, a

lag in the effect of higher demand for goods (new contracts) on demand for labor and, second, a lag between unrealistic demands on the existing workforce and reported labor abuse. With the first type of lag, we expect a natural lag between when companies see an increase in the demand for their goods (*e.g.* due to a lower exchange rate) and the increase in labor needed to satisfy this additional demand. Once higher labor needs are realized, and if labor demand is not met efficiently, we expect this to be manifested in the form of involuntary or unfairly remunerated overtime, coercive targets, or other forms of overwork. With the second type of lag, we expect a lag between when workers are mistreated and when they seek external help. This often arises from poor outcomes following attempts by workers to use an employer's grievance mechanisms to resolve issues, before they seek external assistance. Hence, while it is unclear at what exact lag the effect of labor shortage on worker-reported abuse will be realized, we would not expect our results to be significant when using no lags or too small a lag.

The IV-based estimated coefficients range from 0.082 to 0.187, which corresponds to an additional 8.2% to 18.7% of worker-reported labor abuse if there is a labor shortage increase of one standard deviation in a given week. A labor shortage shock of one standard deviation happens about 10% of the time in the data set, and it happens more often in some provinces (see also Figure 13). Because the baseline rate of abuse in our dataset is 23.7%, such shocks thus cause a 34.5% to 78.8% relative increase in worker-reported abuse over the baseline. Hence, these increases are not only statistically significant but also significant from a human perspective.

We note that the OLS coefficient is not significant and is almost 1,000 times smaller than the IV coefficient. This may be because the IV helped reduce omitted variable bias due to lack of industry controls. We cannot include industry controls in our analysis due to lack of an industry

47

specification in the vast majority of worker voice calls (only 33.8% contain information about a company and/or industry). However, field experience suggests that different industries have different risk factors and means of abuse, due to the nature of work and remoteness of the location. The lack of industry controls therefore creates a bias because some labor shortages are predicted to be associated with higher levels of abuse than others. This can lead the OLS coefficient to be close to zero; *e.g.*, if low-skilled labor shortage in a province is constant over time but comes from a random mix of industries with different likelihoods of abuse, then the OLS coefficient would be zero. In contrast, because the IV is uncorrelated with industry controls, it implements random assignment well.

These findings support the hypothesis that frictions in the local labor market -- as measured by local labor shortages -- contribute to migrant labor abuse. Specifically, unexpected surges in demand for labor that cannot be met efficiently through local labor markets significantly increase reported abuse, which is persistent with a lag of two to four weeks.

Robustness checks

We perform some robustness checks to our results. We study the effect of interest using a subset of the provinces and using different ways of calculating the regression matrix. We also present the result of IV checks, and the result of using a different IV, the US Dollar.

The dataset is messy and underwent heavy cleaning and corrections. Some provinces have a lot more data than others. Some provinces have very few data. It is not clear which provinces we should include and which should be excluded. Even provinces with few data may add important insight to our analysis, because the situation on the ground might be different and worthy of investigation. As a result, we included all provinces with both labor demand and lanor abuse data in our main analysis. However, it is possible that provinces with less data add a lot of

noise to the study results. Hence, we study the same effect of interest among provinces with more data, which are supposed to be more robust.

From table 6, we see that the results from different subsets of provinces are in the same direction, with the coefficient size increasing as we retrict the study to provinces with more data. This is consistent with our hypothesis. Provinces with stronger and more export-oriented economy tend to have more migrant workers, thus it is more likely to receive more hotlines calls from there, whether for general inquiry or to seek help. These provinces are also home to more intense labor shortages from fluctuation in purchasing orders because of their export-oriented economy. As we postulate, the above are exactly the conditions that open the doors for labor abuse, which is why we see bigger exploitative effects among those provinces.

We are also interested in whether the specific parameters of defining and calculating the labor demand shock creates spurious effects. We used the 4 weeks before as the basis to calculate *shock* because we have weekly data and anything shorter than that would be too thin of a basis. Here, we also try 8 weeks. The results are shown in table 7 and show a slightly more prominent effect. Perhaps the previous 8 weeks contitute a better basis with regard to what count as a true labor demand *shock* to employers. Nonetheless, we are reassured to see mostly similar results.

Lastly, we would like to make sure the results are not an artifact of the IV we chose, so we use US Dollars as an alternative and perform the same analysis. United States is the second largest overall trade partner of Thailand. Note that the US accounts for 6.1% of Thai export value, while China accounts for 20% (*24*). As a result, USD is not as good of an IV as the CNY. We transform USD-THB exchange rate the same way as for CNY. Data trend before and after transformation is plotted and shown in figures 13 and 14. IV check results are shown in table 8.

USD is a good IV with 4 weeks of lag, whereas the CNY passes IV tests at 2,3 and 4 weeks of lag. In addition to export volume, the constitution of export by industry can also contribute to CNY being a stronger IV. The top categories of export from Thailand to US tend to require more skilled labor, such as machinery, electrical machinery, vehicles, and optical and medical instruments ([41]). On the other hand, the top categories for China are more oriented toward lower-skilled industries such as rubber, vegetables, raw materials, and intermediate goods ([42]). Because our analysis is focused on exploitation of lower-skilled labor, it is no wonder the CNY is a better IV. Nonetheless, the same analysis results hold for when exchange rates of either currencies pass the IV test (table 9). This gives us more confidence that the results is not merely an artifact of a specific IV choice but reflective of our underlying assumptions.

Table 6. Regression and 2-SLS IV regression results for different regression specifications. IV results are all for labor shortage lag 4 (relative to % abuse), with CNY-THB exchange rate shocks lag 5 (relative to % abuse) as the instrument.

	OLS	2-SLS IV		
	Coeff. $\left(\widehat{\beta_0}\right)$	Coeff. $\left(\widehat{\beta_0}\right)$	Weak Inst.	Wu-Hausman
All provinces	<0.001 (0.002)	0.187*** (0.064)	<0.001	<0.001

Provinces with >= 20 hotline calls	**<0.001** **(0.002)**	**0.285*** **(0.100)**	**<0.001**	**<0.001**	
Provinces with >= 40 hotline calls	**<0.001** **(0.002)**	**0.410** **(0.170)**	**0.006**	**<0.001**	

Outcome: % abuse in a week, on a 0 to 1 scale
Cluster Robust standard errors (province): Yes
Fixed effects: province
IV: CNY-THB, first difference, lagged by one week relative to labor shortage

*, **, and *** denote (two-tailed) significance at the 10%, 5% and 1% level, respectively, for regression coefficient estimates. Standard errors are given between round braces.

Table 7. 2-SLS IV regression results for different lengths of calculating labor shortage shock.

Length of moving-avg. period for calculating shock	4			8		
Labor shortage shock lag	**Est. Coeff.**	**Weak Inst.**	**Wu-Hausman**	**Est. Coeff.**	**Weak Inst.**	**Wu-Hausman**
2	0.082* (0.042)	<0.001	0.04	0.101* (0.053)	<0.001	0.04
3	0.088* (0.045)	<0.001	0.02	0.107* (0.059)	<0.001	0.02
4	0.187*** (0.064)	<0.001	<0.001	0.233** (0.090)	0.001	<0.001

Outcome: % abuse in a week, on a 0 to 1 scale
Cluster Robust standard errors (province): Yes
Fixed effects: province
IV: CNY-THB, first difference, lagged by one week relative to labor shortage

*, **, and *** denote (two-tailed) significance at the 10%, 5% and 1% level, respectively, for regression coefficient estimates. Standard errors are given between round braces.

Table 8. P-value of IV tests for different demand lags and exchange rates of US (second largest export destination of Thailand) and China (largest export destination of Thailand), for the first stage of 2-SLS regression with labor shortage as the instrumented variable and lagged exchange rate as the instrument. No Sargan test conducted because we have the same number of endogenous variables and instruments.

	USD		CNY	
Labor shortage shock lag	**Weak Inst.**	**Wu-Hausman**	**Weak Inst.**	**Wu-Hausman**
1	<0.001	0.76	<0.001	0.20
2	0.001	0.55	<0.001	0.04
3	0.001	0.87	<0.001	0.02
4	0.003	0.003	<0.001	<0.001

Table 9. 2-SLS IV regression results for weekly \% abuse, instrumented by exchange rate shocks with USD and CNY, and significance of corresponding IV tests -- for IVs that pass both the Weak IV and Wu-Hausman tests.

CNY-THB	2-SLS Regression Result		IV Tests	
Labor shortage shock lag	Est. Coeff.	Std. Err.	Weak Inst.	Wu-Hausman
2	0.082*	0.042	<0.001	0.04
3	0.088*	0.045	<0.001	0.02
4	0.187***	0.064	0.001	<0.001
USD-THB				
4	0.140**	0.059	0.003	0.003

Outcome: % abuse in a week, on a 0 to 1 scale
Cluster Robust standard errors (province): Yes
Fixed effects: province
IV: CNY-THB, first difference, lagged by one week relative to labor shortage

*, **, and *** denote (two-tailed) significance at the 10%, 5% and 1% level, respectively, for regression coefficient estimates. Standard errors are given between round braces.
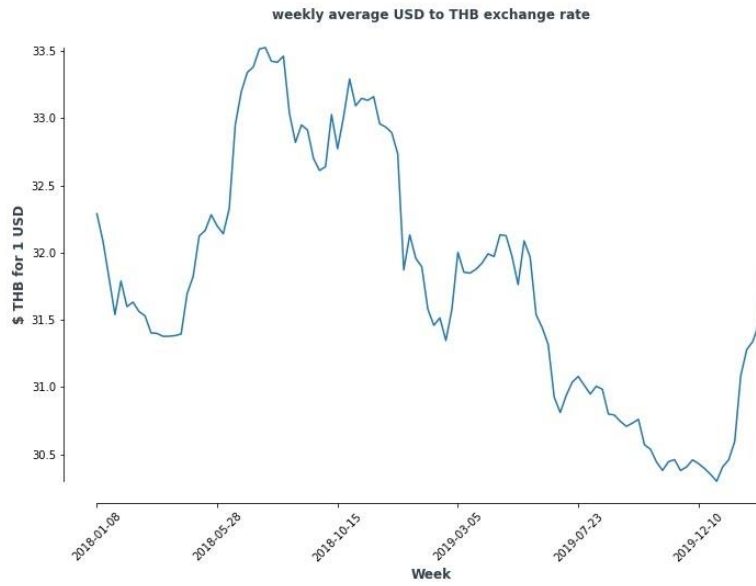
Fig. 13. US Dollar-Thai Baht exchange rate, averaged at weekly level, from Jan. 2018 to Feb. 2020. Source: Bank of Thailand.
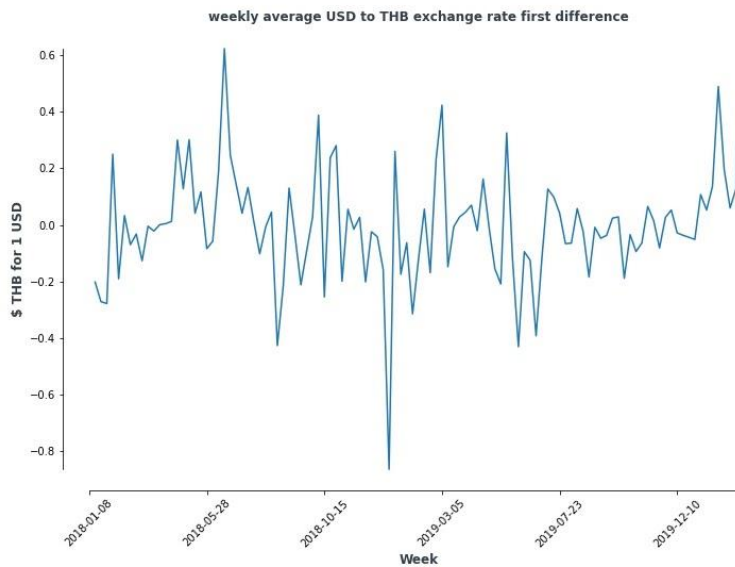
Fig. 14. First difference of Chinese Yuan-Thai Baht exchange rate, averaged at weekly level, Jan. 2018 to Feb. 2020. Source: Bank of Thailand.
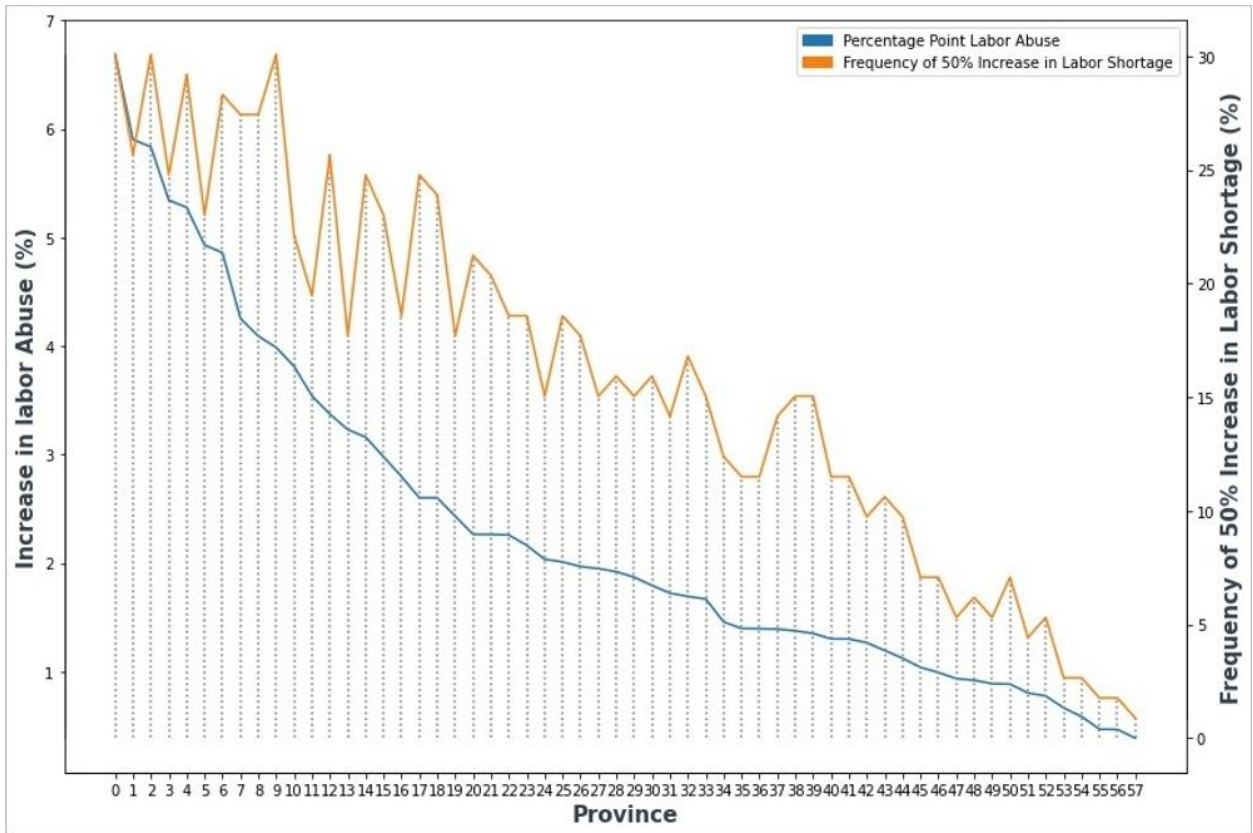
Disaggregation by province

As illustrated by Figures 5 and 9, there is significant heterogeneity in low-skilled labor shortages and worker-reported labor abuse across provinces, *e.g.* due to heterogeneous industry composition and/or remoteness. Labor shortages may thus have a different impact on labor abuse in different provinces. We examine the frequency and estimated impact of a hypothetical labor shortage in each province. Specifically, we examine the predicted impact of a 50% increase in standardized mean labor shortage on labor abuse in each province. Because different provinces have different distributions of low-skilled labor shortages, the standardized mean labor shortage is different in each province and therefore also the predicted impact of a 50% increase.

Figure 15 shows the predicted responses. The left axis shows the predicted absolute increase in worker-reported labor abuse from a 50% increase in standardized mean labor shortage, calculated as:

$$\widehat{\beta_0} \times 0.5 \times \frac{\mu_{S_i}}{\sigma_{S_i}}$$

where μ_{S_i} is the mean labor shortage in province i, σ_{S_i} is the standard deviation of labor shortage in province i, and we use the conservative IV coefficient with province fixed effects in table 4 ($\widehat{\beta_0} = 0.082$). The right axis shows the frequency of a 50% increase in standardized mean labor shortage occurring for a given province in our dataset.

Fig. 15. Estimated increase in labor abuse for a 50% increase in labor shortage, and frequency of such 50% increase in labor shortage occurring, disaggregated by province.

We find that for Bangkok, for example, a labor shortage increase of 50% over the standardized mean in a given week leads to an additional 6.7% predicted worker-reported labor

abuse, and such shortages occur 30.1% of the time. Because the average percentage labor abuse in Bangkok is 17.9%, a 6.7% increase is relatively significant -- such shocks cause a 37.4% increase in worker-reported abuse over the baseline. The provinces with the strongest responses are Bangkok, Samut Sakhon, Chonburi, and Samut Prakan, which are hubs for manufacturing, seafood, and construction in or near the Bangkok megalopolis. The labor markets in these provinces are known to generally be more stressed, as also captured by the high standardized mean labor shortage. These provinces face an additional 6.7%, 5.9%, 5.8% and 5.3% predicted labor abuse if labor shortage were to increase by 50% above the standardized mean in a given week.

We also find correlational evidence that provinces that are generally more stressed (or, alternatively, have a higher average labor shortage relative to the variation in labor shortage) also tend to have shocks in labor shortages more frequently, as can be observed by the parallel trends in Figure 15. This suggests that stressed labor markets are simultaneously more prone to unexpected shortages and abuse.

Chapter 5: Discussion

According to the United Nations Guiding Principles on Business and Human Rights ([43]), businesses have the responsibility to respect the rights of workers and take responsibility for remediating labor abuses within their supply chains. Traditional compliance-driven tools of businesses to identify and mitigate labor risks -- such as audits and social certification schemes -- have proven inadequate ([44], [45]), leading to millions of workers abused and exploited in global supply chains and who are never adequately identified and remediated.

What if we could identify ways to improve business processes that would reduce the amount of harm and exploitation ever experienced by workers in global supply chains, thus reducing rates of human trafficking and forced labor as well as reducing business risk?

This research studies the role that market frictions in low-skilled labor recruitment play in aggravating human trafficking and forced labor outcomes. We find that labor abuse and trafficking peak when mismatches between local supply and demand for low-skilled labor are unexpectedly large, and that such mismatches occur frequently. Using a worker rights-centered labor monitoring mechanism to identify over 2,000 worker-reported instances of labor abuse in Thai factories and worksites where audits and government inspections failed to identify abuses and risks, and an IV approach to analyze this data, our estimates indicate that a shock of one standard deviation in low-skilled labor shortage leads to a 34.5% or higher increase in worker-reported labor abuse in the two to four weeks that follow. Importantly, shocks in labor shortages of such magnitude occur for around 10% of the weeks in the dataset.

These key findings point to the important role for the business and human rights community to develop or adopt processes and technologies that increase safeguarded, credible

worker reporting and reduce time-sensitive supply-demand mismatches in the low-skilled labor supply chain. Five important recommendations for both business and government include:

(i) Global brands and retailers can improve responsible purchasing practices and policies by incorporating a supplier's production capacities (or inferring those from previous production commitments) in their purchasing practices, to anticipate corresponding production pressures on suppliers' workforce and adjust lead times as needed. Previous work has shown that it becomes increasingly possible for global brands to forecast suppliers' production pressures and corresponding likelihoods of unauthorized subcontracting or labor abuse (see, e.g. [20], [21]).

(ii) Global brands and retailers can support and partner with safeguarded, credible worker reporting mechanisms in key sourcing localities, and directly integrate signals from such data into their purchasing practices. Furthermore, data from worker reporting mechanisms and signals about abusive employers should be used to inform migrant workers about forced labor risk at a particular employer *in advance.*

(iii) Suppliers can strengthen HR planning and operations to forecast manpower needs. Invest in developing well functioning grievance mechanisms (particularly inclusive of foreign migrant workers) and engage with independent worker reporting mechanisms to increase visibility of actual working and production practices, as validated by safeguarded workers.

(iv) Governments, in both sending and destination countries, can drive national industry competitiveness and economic opportunities of its labor force through improved systems for regulating the employment of foreign migrant workers. Market frictions can be addressed by (a) reducing reliance on employer sponsorship models that tie foreign migrant workers to single employers; (b) allowing workers more control over where they are employed, with more efficient and transparent in-country movement of legally documented workers between legally registered

employers according to industry and employer demand; and (c) giving workers the right to change employers. The importance of this is applicable beyond the geographies studied in this paper, e.g. see ([27]) for a report on the H-2A program in the United States.

(v) The technology and human rights sector can develop centralized or online platforms for efficient low-skilled labor recruitment and reduced information asymmetries for job seekers.

Together, these actions are expected to significantly reduce frictions in low-skilled labor recruitment processes and therefore incidences of forced labor and other forms of labor exploitation in global supply chains.

Part II:
Integrating market design, causal inference, and contextual bandit algorithms, to strengthen human rights protections in global supply chains

Part II Executive Summary

Employers upstream in global supply chains are often based in the developing world and, due to the absence of centralized labor recruitment platforms, rely on a set of informal and formal brokers to recruit their (often migrant) workers. This process increases the risk of human rights abuses in the workplace due to several factors, including the burdening of recruitment costs onto jobseekers, little transparency and accountability about working conditions, and higher frictions and stress in matching labor demand and labor supply. Current efforts aimed at ending forced labor have focused on creating and enforcing labor standards in the workplace (e.g. through social audits and criminal justice), but none of these address the underlying root cause of informal recruitment, none of these provide real-time monitoring, and all of these are reactive in nature. To tackle these issues at their root cause, we are working with Issara Institute to trial a new approach: the deployment of an online labor platform in Southeast Asia that can facilitate recruitment, provide more transparent job opportunities to workers, help employers recruit more efficiently, and collect data that could extract signals about labor risk in the workplace in real-time. Making this approach a success, and leveraging it to strengthen human rights protections in communities around the world at scale, requires a careful design of the labor platform and the employer-employee matching algorithms in particular. We propose to leverage techniques that combine market design, multi-armed bandits, and causal inference, to devise algorithms that provide optimal matching to workers, learn employer risk scores efficiently, while upholding the highest ethical standards.

Chapter 6: Introduction

In the previous part, we have found that frictions in the low-skilled labor market, in the form of sudden labor shortages, contribute significantly to forced labor. As a result, creating additional standards without also reducing frictions in labor recruitment (e.g. by offering centralized job markets for low-skilled labor) could backfire and worsen the fate of workers.

This prompts us to explore an alternative approach to mitigating forced labor in global supply chains; a centralized matching platform to help low-skilled workers find jobs more easily, and provide more transparency on employer conditions in the process. Job fairs have been shown to improve employment outlook, job mobility, as well as awareness and information about alternatives ([54]). Reducing search cost not only benefits job seekers, but also improves labor allocation efficiency ([55]). What if we can create a perennial job fair in the form of a digital labor market? We propose to leverage the combination of market design, causal inference, and reinforcement learning to make sure our inference on employer conditions is reliable, and that we continuously improve and act on our inference to help workers find better jobs.

A key contribution of this work will be to develop a matching algorithm (to match job seekers and employers) and a causal inference algorithm (to infer labor risk at an employer) that work in tandem. The matching algorithm facilitates *stable* job matching based on preferences from both workers and employers, while the inference algorithm uses worker behavioral data to infer *red flags* on employers that may indicate abusive situations. The matching algorithm helps with inference by mimicking conditions similar to random controlled trials without sacrificing worker choice and preferences; critically, this is possible because the platform would be dedicated to relatively low-skilled workers, who often have many overlapping skills and preferences. The inference algorithm helps with matching by continuously updating the *red*

flags, leveraging tools from the contextual bandits literature. It is important to note that, as a result of this combined approach, we would never knowingly or actively perform any randomized trial with workers. The ethics of conducting randomized trials in high-stake situations that potentially involve forced labor can be questionable, and we avoid doing that.

This work will have three desirable outcomes, 1) ability to continuously and automatically detect "red flags" based on data, which can inform regulatory and compliance efforts in forced labor and human rights; 2) reducing forced labor risk due to better and more transparent "red flags" on employers in global supply chains; 3) reducing forced labor risk due to reduced demand-supply mismatches.

The Golden Dreams market platform

This work will be more impactful when deployed in a real job matching platform where it needs to be to reduce labor abuse. Luckily, our partner, Issara Institute, launched its new version of the Golden Dreams recruitment marketplace in August 2021 in Thailand ([46], [47]). It is a mobile application that aims to help workers find more good jobs and end the abuse, exploitation, and human trafficking from the international labor recruitment process. It is in workers' native language and developed through extensive user research and based on Issara's field expertise. Its first version, released in 2017, is a Yelp-like platform that allows local workers to provide reviews about recruiters and employers. The second version, which was released in 2021, comes with a secure job platform that directly connects employers facing labor shortages with job seekers.

Golden Dreams was designed to address some of the biggest drivers of labor exploitation. In job advertisements, employers must disclose information about salary and overtime rates, working hours, benefits, job description, requirement, and conditions, recruitment cost,

accommodation, grievance mechanism, and the exact location. Hidden or high recruitment cost has been identified as one of the biggest contributors to labor exploitation in Southeast Asia, which is home to a huge number of migrant workers. Moreover, the platform, along with Issara's Worker Voice help hotline, empowers workers to voice their opinions and drive change from the ground up, an approach complimentary to the traditional top-down approach of auditing and international intervention.

The platform can be made better in several areas. First, at the time of writing, the platform has a decentralized matching approach, where job seekers contact each employer or recruiter separately and then compare the different options they receive. If they receive more than one job offer, they will need to compare and turn down all but one. Employers face the uncertainty of job offers not taken and will need to make additional offers to fill those remaining vacancies. Job seekers may then receive new offers, and they can be better than those they received earlier. Job seekers may turn down previously taken offers and create further cascading effects, or they may not be able to do that and regret their earlier decisions. In any case, the situation can create significant uncertainty frustration, and regret among both job seekers and employers.

Second, while it is empowering to let workers directly voice their opinions, those opinions may not be frequent or comprehensive enough. People are more likely to write reviews when they receive extraordinarily good or bad services or treatments. Thus, we can miss the true opinions of the bulk of people, the silent majority. On the other hand, we can potentially have a more accurate picture of employer and employee satisfaction and help people make better choices if we can capture behavioral information automatically and learn about good or bad employers from that on an ongoing basis. Thus, everyone would contribute to the pool of

opinions through their day-to-day behaviors. This approach is not only more timely but also more reflective of the entire population.

Additionally, there can potentially be coercive or otherwise untruthful reviews that jeopardize people's trust in the platform reviews. When reviews matter, people find ways to hack them to their own advantage (see, e.g., Amazon [48] and Yelp [49]). When reviews are public and non-anonymous, employers can force workers to write positive reviews or remove negative reviews. When reviews are anonymous, malicious users can write fake reviews for revenge or competition. We can mitigate this problem by using behaviors that are hard to hack as proxies to come up with implicit scores to complement explicitly written reviews.

Together, we envision the next generation of the Golden Dreams platform that is best equipped to help workers find more good jobs and help employers meet their labor demands and remove labor exploitation in the system. The process will be smooth, straightforward, and reliable for all parties involved. It will encourage trustworthy behavior in recruitment through information gathering, inference, and disclosure. Lastly, it provides additional information and tools to regulators, NGOs, and corporate citizens in the supply chain to enforce anti-exploitation and anti-trafficking policies. More on that will follow in the discussion section.

Theoretical contribution: learning and action in a complex and resource-constrained world

Automated decision-making based on data penetrates into more and more aspects of modern human society. It transforms our lives by facilitating communication, learning, decisions, and transactions with unprecedented scale and speed. We think there are four common pillars to this paradigm. Their relative importances differ based on the usecase; different combinations of them give rise to different solutions to those usecases.

1) **Perception, planning, and control**

Perception, planning, and control are the defining characteristics of modern robotic and autonomous systems. Perception means the system can collect information about the state of the world in order to make decisions and achieve its goals. Planning means the system can make sense of the gathered information and devise a course of actions. Control means that it can implement those decisions in a way that influences the world to achieve its goals. Since the goal of control is to change the state of the world, it necessitates a new round of perception, planning, and control to take the changes into account. This process goes on continuously. A defining characteristics of this kind of problem is that previous actions affect the probablistic distribution of information collected sebsequently. Examples of such system include most robots and autonomous vehicles.

2) Uncertainty, learning, and adaptation

The world is inhenrently complex and uncertain. An agent with a goal might need to perform certain actions to uncover previously hidden information, analyze the information collected, and adjust its rules, decisions, or behaviors. One such example is recommendation systems, which collect and analyze users' behavioral data to learn their preferences and recommend items that are more likely to lead to desirable outcomes, such as interactions for Facebook timeline ([50]) or clicks for Google Ads ([51]). Intelligent robotic systems are increasingly moving towards achieving this goal ([52]), thus combining 1) and 2). Multi-armed bandits is another classical problem that combines continuous action (planning and control) and feedback (perception) as well as uncertainty (in reward).

3) Resource constraint

While digital goods usually can be multiplied with minimal cost, most things are not, such as physical goods, services that require human, or time. Even though automated decisions are fast

and scalable, resource constraint underpins much of practical applications of such systems. It takes time, resources, or other consequences to gather data and learn. Traditional optimization problem hinges on resource constraint as its defining characteristics.

4) Decentralized information, preferences, and needs

Automated decision-making systems are often created to consolidate otherwise decentralized information, preferenes, and needs in order to make efficient or welfare-maximising decisions. If everyone's preferences were the same, there would not be any decision to be made – it won't make any difference anyways. On the other hand, because information is decentralized, automated decisions always need to work with individuals to collect information and account for it. Hayek argues that, for the same reason, a central authority would not be optimal ([53]). Automated decision-making systems are often similar to centralized authorities. Recommendation systems try to collect those information and make decisions for people. The study of market design arises when a problem combine resource constraint and decentralized preferences. If schools have unlimited seats, we can just assign all students to their favorite school.

Complex problems in the real world, however, often combine all four characteristics. Our job marketplace is just such a case. We have workers with different preferences and needs (4). We have limited number of jobs from each employer (3). Both workers and employers need to gradually learn about what kind of match best suits their needs even when they already know their own preferences because some information are strategically witheld, such as likelihood of labor exploitation (2). Lastly, workers and employers need to act while learning, and do it iteratively with the goal of optimizing their long-term welfare (1). As we will see in the next section, there are work that tackle combinations of some of these characteristics, but there are

few work that formulate their problems with all four, perhaps for good reasons. It turns out to be quite challenging.
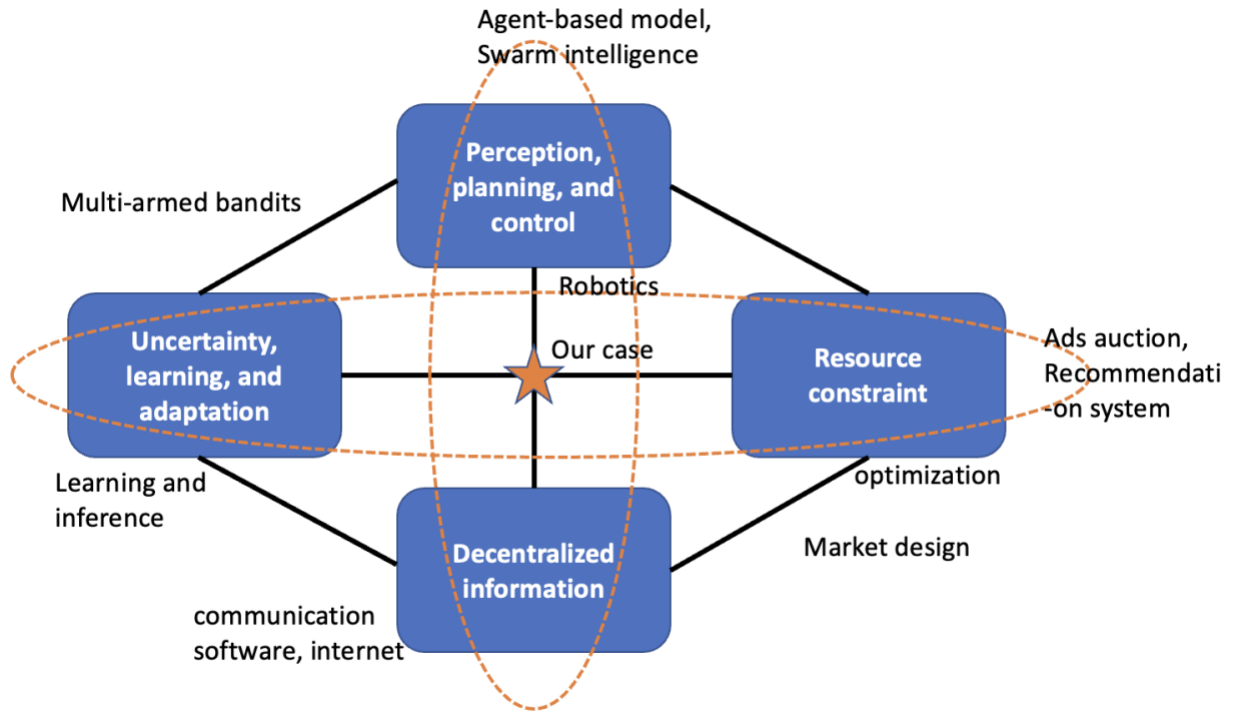

Fig. 16. An illustration of the four common pillars in automated decision-making systems

Chapter 7: Literature Review

We primarily draw on two bodies of literature, that of market design and causal inference, and that of multi-armed bandits algorithms.

Market design and causal inference

The market design literature has been evolving in the past several decades and becoming a rich body of insights. One of the most famous algorithms used in matching two sides of the market that produces *stable* matchings is called Deferred Acceptance (DA), which was first introduced and popularized by Gale and Shapley (1962) ([56]). It has been used in one-to-one and many-to-one matchings, such as the labor market of medical interns and hospitals (Roth 1984 [57]), and students applying to schools (Abdulkadiroglu *et. al.* 2006 [58]). Take the example of students applying to colleges. We have a list of colleges and a group of students. The colleges each have a capacity constraint on how many students they can accept. Both colleges and students have preferences over the opposite side. The model setup optimizes for matching colleges and students to their most preferred choices as much as possible under various constraints that ensure the market functions properly. Getting into one's favorite school is better than getting into the second favorite.

Often, in such market setups, we would like to estimate causal effects of certain features or outcomes, such as enrollment in a certain kind of school on educational outcome. Such estimations need to consider not only features of market participants, but also their preferences. This is because getting into one's preferred school likely has a positive effect on educational outcome compared to getting into a less preferred one, even when the two schools are almost identical when measured objectively. Abdulkadiroglu, Angrist, Narita, and Pathak (2017) [59]

described such a method to efficiently estimate causal effects using propensity scores derived from random tie-breakers due to preferences and capacity constraint.

The Deferred Acceptance (DA) algorithm

Here we describe the Gale-Shapley algorithm, also known as Deferred Acceptance (DA) algorithm, in more detail. It is a seminal algorithm first described by David Gale and Lloyd Shapley ([56]) that has been subsequently adopted in numerous real-world applications, most notably labor clearing houses and school choice systems.

This procedure assumes that there are two sides of the market with preferences over the other side. Let's call them students and colleges. We usually assume there are no ties in preference lists for simplicity, but ties can be broken randomly or according to certain rules if necessary. One side, say the students, apply to their favorite colleges. Some colleges may receive more applications than what they can accept, in which case they reject their least favorite applicants and temporarily hold the rest. Colleges that do not exceed their capacity hold all their applicants. Rejected applicants may not apply to those colleges that already rejected them, but they then proceed to apply to their favorite ones among the rest. Colleges that receive new applications consider all applicants together and reject their least favorite ones until they are at capacity, and temporarily hold the rest. The process goes on until there are no more new applications, at which point all colleges accept all applicants they are holding. The process is bound to terminate in finite rounds, yielding a final, non-empty set of matches.

Gale and Shapley prove that this resulting match must exist and is *stable* and *optimal* for the applicant side. An *unstable* match happens when there are two participants, college A and student 1, who prefer each other more than whom they are assigned. *Optimality* is defined among the set of *stable* matches.

Markets and multi-armed bandits

One limitation when applying the framework of Abdulkadiroglu *et. al.* in our scenario is that they assume known preferences. As a matter of fact, most literature on stable matchings assume preferences that are known or can be discovered through a few interactions before the matching (Liu *et. al.* 2020 [60]). In reality, people make choices based on incomplete information about the others and themselves, and they learn more about their options as well as their own preferences iteratively. Such is the mantra of the multi-armed bandits (MAB) literature. It is a study of making decisions under incomplete information and continuously improving one's knowledge and making better choices. Both parts, the learning as well as the continuous improvement in subsequent choices, are salient to our application in a labor market.

Liu, Mania, and Jordan ([60]) introduce a model setup that combines centralized markets and bandits. In their model, a group of players (e.g. companies with tasks) pull a set of arms (e.g. workers who work on those tasks) to get stochastic rewards (e.g. task performance) drawn from 1-sub-Gaussian distributions with means determined by the matching pair. These means are not known beforehand and must be learned through repeated interaction. If multiple players ranked the same arm, a conflict arises. Only one of the players get to pull the arm as decided by that arm's preferences. The rest attempt to pull their less favorite arms as decribed in the DA algorithm. Based on this formulation, they then develop a notion of *stable regret* (the difference between what reward a player end up getting versus what it can reasonably expect under stable machings) and study its convergence rate with the goal of minimizing the regret. They explored two algorithms for this tasks, Explore-then-Commit (ETC) and a centralized version of the well-known upper confidence bound (UCB) algorithm.

This is an innovative line of work that is relevant to our problem. We would like to see more developments in two areas, 1) information sharing between players and how it can improve the learning and matching outcome, and 2) to avoid the exploration phase in the learning strategy due to the high moral stakes of exploration that potentially involve forced labor.

Contextual bandits and the exploration exploitation tradeoff

Both learning algorithms in [60] reflect another central theme in the MAB literature, that of explore and exploit tradeoff (see, e.g. [61]). In ETC, players are forced to explore new options for a predetermined number of rounds to learn about their expected reward, and then exploit the best option(s) after that. In UCB, the process is more adaptive and players gradually explore less as they learn more about their options and build up certainty. The exploration is incentivized through imposing on players an intentionally optimistic view of the world.

Exploration is a prized activity in the MAB literature. Frazier et. al. ([62]) study incentives to encourage exploration by myopic and selfish players and the total reward achieved by the system. They provide a characterization of the relationship between the two – higher incentive payment is needed to achieve higher reward. Later on, Chen et. al. ([63]) expand this research to study what would happen if those players are heterogeneous and know rewards given by each arm to previous pulls by other players. They find that such heterogeneity and information sharing provided "free exploration" – the regret bound is contant as opposed to growing logarithmically in standard settings, which is a big improvement.

Still, can we avoid incentivizing players at all? Active exploration can take the form of information disclosures or recommendations ([60], [64]) that do not represent immediate best interests for participants, or monetary subsidies for certain choices ([63], [65]). It can be ethically

controversial in high stake situations such as (potentially abusive) employment, even if it satisfies incentive compatibility and individual rationality constraints.

Bastani *et. al.* ([66]) explore such a possibility because "exploration may be costly or unethical". They show that when there are two arms and players are sufficiently diverse in their preferences, exploration is not necessary and a greedy algorithm is optimal. They then expand to the more general setting to provide a "greedy-first" algorithm that performs greedy exploitation first and checks whether the system is learning the true parameters at a suitable rate, and switch to forced exploration if is it not learning. This result is encouraging, but limited to the single player case. To be able to apply their results to our setting, we need to extend their result to a market setting where there can be multiple players in each round and thus collision of choices.

Chapter 8: Model

Our technical approach has two main components, the matching algorithm, and the inference algorithm.

Model setup

Matching

We model employers as bandit arms, and workers who apply to jobs as players who pull the arms. The market has K arms and an unknown time horizon of T time steps. At time $t \in [T]$, observe M_t **new** players, which is a set $[M_t]$, and each of them $p_{j,t} \in [M_t]$ has a d dimensional context vector $X_{j,t} \in R^d$. For simplicity, we assume the number of players each round is equal, i.e., $\forall t$, $M_t = M$, and that the demographics don't change ($\forall j, t$, $X_{j,t}$ is drawn *i.i.d.* from some unknown distribution $P_X(X)$). For simplicity, we also assume each employer has one job in each round, which correspond to the MAB setup that each arm can be pulled by one player at a time. If an employer advertises and fills multiple job vacancies in certain rounds, we collect more data about it and make better inference. So our assumption almost gives a theoretical lower bound. Because data suggests that, in practice, there are usually significant gaps in unmet labor demand, there probably tend to be more jobs than workers. In a worker proposing DA algorithm, the job vancancies that don't get filled will never get filled in any possible scenario in that round (rural hospital theorem, Roth 1986 [67]). Therefore, we drop those extra arms and assume the number of arms and players in each round are equal ($M = K$). Workers that didn't get their preferred job go to their next preferred job until they get the best available job, which is guaranteed by the DA algorithm. The DA algorithm is treated as a blackbox for now, and will be described in detail later.

Arm $a_i \in [K]$ has $\beta_i \in R^d$ as its feature, which include attributes that need to be learned such as its abusiveness. For simplicity, we assume context vectors always have an extra 1 appended to incorporate fixed effects for arms. These arms are employers that do not change from round to round. Players do not know $\{\beta_i\}_{i=1}^K$ in advance. Arm a_i also has preference over players in round t in the form of $\prec_{i,t}$ and associated preference ranking $pr_{i,t}$ (most to least preferred), which is determined by preference function f_i of each arm that maps context vectors from round t into desirability scores. If arm a_i prefers player $p_{j,t}$ over $p_{j',t}$, we say $p_{j',t} \prec_i p_{j,t}$. Similarly, we denote the expressed preferences of players based on assumed arm parameters by $\widehat{\prec}_{j,t}$, and their true preference as $\prec_{j,t}$. It should be clear from the context whether we are using the $\prec$ sign to denote player or arm preferences. We use $r_{j,t}$ and $\widehat{r}_{j,t}$ to denote the list of rankings of arms by player $p_{j,t}$ based on the true or estimated arm parameters. We assume all players and arms are acceptable to each other. When player $p_{j,t}$ from round t successfully pulls arm a_i as determined by a DA matching process, we call it a match and denote the result by $m_t(j,t) = a_i$. The associated reward $Y_{i,j,t} \equiv X_{j,t}^T \beta_i + \varepsilon_{i,j,t}$ comprises of the product of the player and arm pair and an independent σ-subgaussian random variable. All matches in round t is denoted by m_t.

The "reward" for pulling an arm can be interpreted as job satisfaction, which comprises both workers' preference over objective features of the job, such as salary and hours, and unexpected deviations from objective features, which we attribute to employer risks. At each step, we try to infer risks, which is described in more detail in the next subsection ("Inference"). Job satisfaction can be approximately measured by job turnover time (the quicker it it, the lower satisfaction) or frequency of searching and applying for the next job. They are not perfect measures, but they have the benefit of being continuously measured, as opposed to job

satisfaction surveys, which can only be conducted periodically and also suffer from internal biases of responses and selection bias of who responds.

Standard bandits literature (*e.g.* [66]) usually assume an *oracle* policy π^* that, unlike players, knows $\{\beta_i\}_{i=1}^K$ in advance. Thus, given preference context vector $X_{j,t}$ of player $p_{j,t}$, π^* is able to find the true ranking of arms for that player. Player $p_{j,t}$, on the other hand, submits its preference based on its estimates at round t of arm parameters, $\{\widehat{\beta_{i,t}}\}_{i=1}^K$. It then receives reward from its submitted preference and the resulting match, which is based on the true arm parameter instead of the estimated parameter. The badits literature usually assumes the *regret* as the difference between the arm chosen by a player and the best arm chosen by the *oracle*, and tries to minimize the cumulative expected *regret*. This assumption, however, does not consider the possibility of collision of rankings in a marketplace, where the outcome not only depends on one's preference, but also the preferences of others that may be misguided.

Similar to [60], we use the concepts of *valid match*, *stable match*, *optimal match*, and *pessimal match* to create meaningful benchmarks under this circumstance. Stable match is a standard concept in market design literature, meaning there does not exist a pair from two sides of the market that both prefer each other than their current matches. There can be multiple sets of stable matches in a market depending on randomness or the matching algorithm, even when everyone have fixed preferences. Player $p_{j,t}$ and arm a_i as valid matches if they are matches in one of the stable matches. The *optimal* ($\overline{m}(j,t)$) and *pessimal* ($\underline{m}(j,t)$) matches, consequently, are the most and least preferred ones among one's valid matches. We use *optimal* and *pessimal* matches instead of the oracle as the aspirational benchmark, and define *regret* as the difference between these benchmarks and what is actually realized. Thus, player-pessimal stable regret

$$\underline{r_{j,t}} \equiv X_{j,t}^T \beta_{\underline{m}(j,t)} - X_{j,t}^T \beta_{m_t(j,t)}$$

76

while the player-optimal stable regret

$$\overline{r_{j,t}} \equiv X_{j,t}{}^T \beta_{\overline{m}(j,t)} - X_{j,t}{}^T \beta_{m_t(j,t)}$$

In subsequent rounds of the matching algorithm, a new set of workers apply to jobs from the same pool of employers following the same process. They can consider and may benefit from inferred employer risk scores from previous rounds, but those scores have uncertainty. The contextual bandits literature helps us understand how this kind of information disclosure affects subsequent behaviors and our ability to learn further about risk scores. For any single player, we seek to minimize the cumulative expected regret $\underline{R_j} \equiv \sum_{t=1}^{T} \underline{r_{j,t}}$ (player-pessimal) or $\overline{R_j} \equiv \sum_{t=1}^{T} \overline{r_{j,t}}$ (player-optimal).

Inference

We first define $X_t \in R^{M \times d}$ as all context vectors from all players from period t, $X \in R^{M \times T \times d}$ as all context vectors from every period in T, $Y_i \in R^{M \times T}$ as all outcomes for arm i, and $Y \in R^{M \times K \times T}$ as all outcomes for all arms. Note that Y_i and Y can have many empty entries (M-1 empty ones out of every M) since only one player gets to pull any arm in a given round. Similar to Bastani, for any time $t \in [T]$, we let $S_{i,t} \equiv \{j, t' \mid \pi_{j,t' \leq t} = i\}$ denote the set of player j and time t' pairs that pulled arm i within the first t rounds. We use it to subset the outcome Y and corresponding context vectors X of players who are responsible for that outcome. Thus,

$$Y(S_{i,t}) = X(S_{i,t})^T \beta + \epsilon(S_{i,t})$$

Arm parameters $\widehat{\beta_i}(S_{i,t})$ is estimated for arm i based on results in set $S_{i,t}$ with the standard **OLS estimator**.

$$\hat{\beta}(S_{i,t}) = (X(S_{i,t})^T X(S_{i,t}))^{-1} X(S_{i,t})^T Y(S_{i,t})$$

The estimated $\hat{\beta}$ would include factors associated with employer risk and abusiveness. We would like to make robust and reliable inference on employer risk scores based on objective features (such as wage, location, and industry), preferences, and outcomes. Classical causal inference methods are not able to elicit or make use of preference information well, and thus can produce misleading results. For example, workers don't always get to work at their most preferred job due to limited vacancies, which can impact the outcome and erroneously be attributed to other factors. The impact here is an omitted-variable bias where the omitted variable, preference, positively correlates with job satisfaction. Traditionally, researchers have used natural or artificial randomized experiments to eradicate the bias coming from subjective preferences, but it would be ethically highly controversial to do that in our setting. We address the omitted-variable bias using a propensity score based approach first pioneered by Rosenbaum and Rubin ([68]) and later extended to the market setting by Abdulkadiroglu *et. al.* ([59]).

Following Abdulkadiroglu *et. al.* ([59]), we partition the model and data by the arm in a one-versus-rest fashion. We are interested in the effect of being matched to arm *i*, and assume the number of non-compliers, i.e. workers who are matched to a certain job based on their preference who end up not taking the job, is negligible. The key value in this approach is the propensity score $p_D(\theta)$, which encapsulates preferences of both workers (players) and employers (arms) over each other. It can be estimated using the analytical or simulated approach described in [59]. The details are omitted because we do not modify the approach. More details on how we implement it will follow in subsequent simulation work.

Again, we subset and only look at players in $S_{i,t}$ for different arms *i* at different times. $d_i(x)$'s are dummies indicating values of their estimated propensity scores of being matched to arm *i*, $\widehat{p_{D_i}}(\theta)$, indexed by *x*; and $\alpha_i(x)$ are the associated "score effects". We also include

controls for players, $X(S_{i,t})$. $\alpha_i(x)$ and β_i are coefficients associated with arm i, with β_i being

the coefficient of interest. Note that β_i also includes fixed effect for this arm because of the extra

1s in $X(S_{i,t})$. Thus, for all players in $S_{i,t}$ that are matched to arm i,

$$Y(S_{i,t}) = \sum_x \alpha_i(x)d_i(x) + X(S_{i,t})^T\beta_i + v_i$$

With that, we will be able to use OLS to estimate β_i while taking into consideration the

expressed preferences. We will refer to this estimation as the propensity score (PS) estimator for

the rest of this text. It can also be subsetted by results before round t to derive ongoing best

estimates $\widehat{\beta_t}(S_{i,t})$.

We use objective behaviors observed in the platform, such as turnover rate and job

searches, as proxies for the outcome, i.e. job satisfaction, based on the assumption that,

everything else being equal, workers in abusive situations will search for other jobs more

frequently and change jobs more quickly. These proxies will be inverse to reflect that lower

turnover means higher job satisfaction. Each employer will get a score from our inference

algorithm based on its turnover and behaviors of its employees. We will also incorporate worker

reviews of employers and feedback from flash surveys, but we are aware that review data can be

misleading at times, such as if workers are under coercion to write good reviews, or face

retribution for writing bad ones. Discrepancies between reviews and inferred scores can also be

highly indicative of abusive employers. These scores and discrepancies will form the basis of our

"red flag" system.

A greedy algorithm – free exploration from market competition

We look at the following algorithm.

Centralized greedy algorithm

Initialize $\hat{\beta}(S_{i,0}) = 0 \in \mathbb{R}^d$ for all arm $i \in [K]$

For $t \in [T]$, **do**

    For $j \in [M]$

        Observe $X_{j,t} \sim P_X$

        Sort arms into preference rankings $\hat{r}_{j,t}$ according to

        $X_{j,t}^T \hat{\beta}(S_{i,t-1})$ (descending, break ties randomly)

    End for

    Arms determine $pr_{i,t}$ - their preference of players in this round

    Gale-Shapley platform computes m_t - all the round t matches

    For all j, **Update** $S_{m_t(j,t),t} \leftarrow S_{m_t(j,t),t-1} \cup \{(j,t)\}$

    Calculate $\hat{\beta}(S_{i,t})$ for all arms i based on OLS estimator or

Propensity Score estimator

End for

The OLS and Propensity Score estimators are as in the inference section. Note that until we have

d samples, OLS estimator is not well-defined. The PS estimator faces a similar issue depending

on the number of brackets in x. We do not calculate or update $\hat{\beta}_t(S_{i,t})$ in those situations until

there is enough data. Overall, the difference with [60] is that 1) we use best (most likely)

estimate instead of an optimistic upper confidence bound (UCB) to guide action, thus removing

the ethical dilemma of potentially misleading exploration induced by UCB; 2) we model players

with context so that we can leverage all outcome observations from different players for the

same arm to estimate arm parameters (information sharing); 3) our estimation arm parameters

can incorporate player preferences, resulting in less biased results. The Centralized Explore-

Then-Commit (ETC) algorithm in Liu et. al. also uses the expected reward (mean) instead of the

optimistic one to guide action, but it forces players to explore nonetheless. Compared to the

model in [66], we introduced a less biased estimatior that incoporates player preferences, and

turned it into a centralized matching market. We will now proceed to study and prove the

learning rate and cumulative regret of this algorithm. It is our assumption that such an algorithm,

which does not intentionally incentivize exploration, nonetheless is able to explore different

options to learn about them because of the collision of wants from market competition.

Next steps

As in standard bandits literature, we are interested in proving the regret bounds of this

algorithm and show its performance through simulation or real data. By the time of my

graduation, I have made several attempts to prove the regret bounds, but have faced several

challenges and have not completed it.

Concretely, one of the attempts was inspired by [66] to show that 1) every arm will be

explored with sufficient probability despite the lack of forced exploration (lemma 3); 2) the

estimated arm parameters will converge to their true values (lemma 5) since the minimum

eigenvalue of the sample covariance matrix is positive and bounded from below (lemma 4); 3)

the instantaneous regret bound in each round shrinks as a function of $\frac{1}{t-1}$ with respect to time t,

thus making the cumulative regret $O(\log(T))$. The first step is still guaranteed since every

acceptable arm will be explored by some players since not everyone can get their favorite. The

critical difference in our new model is that, exactly because we have a market and thus collision

of preferences, the matching outcome is highly chaotic and depends on the estimation error of

arm parameters of not only oneself, but also those of others. Moreover, every player's

subsequent ranking lists are also affected by earlier outcomes that are hard to analyze, which

makes the sample distribution for each arm not i.i.d. Bastani et. al. were able to deal with the

non-i.i.d. issue by breaking down the scenarios depending on which arm has higher estimated reward based on estimated arm parameters. But in our case, that is not enough since players are not guaranteed to get what they prefer, further complicating the analysis. Thus, the matching outcome and player-optimal or player-pessimal regret becomes hard to analyze.

The proof of regret bound of centralized ETC algorithm in [60] offers additional interesting insight. Their proof consist of 1) showing players incur player-optimal regret if and only if they make mistakes on arms more preferrable than their optimal stable arm $\overline{m}(j,t)$, i.e. ranking an inferior arm that $\overline{m}(j,t)$ before it; 2) the probability of such mistakes shrinks exponentially as the number of rounds of forced exploration h goes up; 3) since the regret in forced exploration phase grows linearly with h, the total regret can be bounded logarithmically with regard to the total number of rounds T by choosing h as a logrithm of T. The key here is that the chance of mistakes shrinks fast and evenly because forced uniform exploration guarantees i.i.d. data. Our case does not force any exploration and does not provide similar guarantees.

However, one can argue that centralized ETC is a very useful simplification to our model that is able to provide regret bounds. Centralized ETC can be thought of as a simplification of our model under the following reasoning. Firstly, before we gather enough data to make the first inference (d samples for OLS estimator), players in effect explore blindly with regard to the hidden attributes such as abusiveness. As the market grows bigger, the exploration is approximately uniform across all choices similar to ETC. The difference is that we do not externally impose the number of rounds of uniform exploration. Second, to make the proof more straighforward, they estimate expected rewards using an empirical mean for each arm and player pairs, which is a simplistic regression that ignores all player context information. Retaining those helpful information only help make better inference.

Thus, for next steps, we plan to build empirical simulation models to further study the behavior of our model setup and compare it to benchmarks mentioned in this work or otherwise. I refer interested readers to the following code repository (credit and huge thanks to my collaborators Ravi B. Sojitra, PhD student in Management Science & Engineering at Stanford University, and Yuan Shi, PhD Student at MIT Operations Research Center) (https://bit.ly/job_matching_and_inference) for subsequent ongoing work.

Chapter 9: Putting theory into practice

There are practical considerations when deploying this or similar matching algorithms into the Golden Dreams platform.

Preference orders

There can be thousands of jobs and employers at any point on the platform that job seekers must choose from. They can quickly become overwhelming. Therefore, we will need a way to help job seekers express their preference accurately to get jobs their like. This can take the form of ranking jobs for them based on revealed or inferred preference weights over job attributes. Some people might focus on salary, some on working hours and accommodation, and others on location.

Field interviews suggest that job seekers on the Golden Dreams platform generally don't feel comfortable making abstract generalizations about their preferences. It is indeed difficult to express one's preference in precise numbers. Thus, it might be easier to infer their preference from real or synthetic jobs. For example, we can imagine a pre-assessment with an user interface similar to dating applications, where workers are shown one or two jobs with different attributes at a time and asked to either like or dislike them. That will give us a crude estimate of worker preferences. We will then use this knowledge to rank real jobs for them, and then allow them to adjust the order list. We then update our inferred preference weights from the adjustment and iterate on this process.

Strategic behaviors

One of the most important principles of market design is to make sure that it is straightforward – strategic behaviors aimed at gaming the system is discouraged. It is usually achieved by making sure truthfulness is the (weakly) dominant strategy and always achieve the

best outcome (at least never worse than otherwise). Strategic behaviors in this market can include:

1) Workers misrepresent their preferences.

2) Employers misrepresent their preferences.

3) Workers write untruthful reviews on their own or under coercion.

4) Workers or employers hold off until the next round.

5) Workers and employers go off the platform to find matches.

6) Employers hire people to impersonate job seekers to manipulate their ratings or review.

7) Workers act collectively to make certain demands.

8) Employers act collectively on some agreement.

Among these possibilities, our matching algorithm design can mostly preclude 1), 2), and 5) from happening. It has to do with its property on stability and strategy proofness. Stability guarantees that there is no pair of workers and employers who prefer each other than their original matches, so it doesn't improve their outcomes to try to make separate arrangements off the platform. Strategy proof means that the best strategy for participants is honesty, because they cannot achieve better results from strategizing. This is guaranteed for workers by DA. There is no matching algorithm that yield stable results and is completely strategy-proof for both sides of the market ([69]). Yet it is almost strategy-proof because, as shown by Roth and Peranson, the number of different outcomes shrinks quickly as the market size grows and becomes negligible over 100 participants[1] ([70]). It is also almost the best we can do, as no algorithms that

---

[1] When there are 100 workers and 100 jobs/employers, and each rank 10 top choices, the number of workers that can be matched to different jobs is expected to be less than 3 according to simulation, leaving very little room for strategic or deceptive behaviors.

implement stable matching achieves obvious strategy-proofness, a stronger requirement over strategy-proofness ([71]).

We think that 4) and 7) are probably reasonable behaviors. It is costly for workers or employers to hold off, as workers would be earning less income for the time missed, and employers would delay their production. If they still choose to hold off, it would be a true reflection of their preference that their current options are not acceptable. Collective action or union formation is a strategy commonly encouraged by international organizations and labor rights advocate ([30]), and we concur with this common practice.

The other possibilities can be a little problematic. Based on social norms and international standards, employers are encouraged to cooperate to do certain things (such as labor condition improvements) but not others (colluding to suppress wages, preventing certain members from participating in the market, etc.). This kind of collusion has been considered by some as the most important practical considerations in designing a good market ([72]). Property of strategy-proofness in the DA algorithm does not prevent collusion on wage suppression. Because in DA, reward from matching is assumed to be purely from the match one gets, yet in a practical labor market, reward of employers is the difference between the value of the match (worker) and the wage one pays. If every employer agrees on a ceiling on wages, they all benefit from this collusion and every worker suffer losses.

There are a few ways to potentially mitigate this problem. One can expect collusion to be increasingly difficult when there are more employers. Thus, given we don't compromise selection criteria and background check of employers, we can make it as streamlined as possible to join the platform. We can also periodically conduct surveys and interviews with market participants to learn whether they think the other side is colluding or breaking platform rules, ot

learn such feedbacks from grievance mechanisms. We might also be able to detect unusual activities from platform key performance indicators, such as the average wage, turnover rate, time to offer, etc.

Untruthful behaviors such as 3) or 6) are also challenging and pose risks to the success of Golden Dreams. This is one of the motivating factors for us to develop a simultaneous matching and inference process using data that is hard to compromise, such as frequencies of job search, interviews, and turnover rate. These activities are costly to fake, and we think they are closely related to job satisfaction, which in turn reflects work conditions. Differences between our inferred risk score and explicitly written reviews can be normal and informative, but huge gaps will be investigated. We think such investigations will prevent and rectify untruthful behaviors.

Do we have effective and fair proxies to labor abuse?

This leads us to the next consideration, which is to prove the proxies we use are indeed indicative of job satisfaction. We can empirically test this hypothesis. We will need to capture data about workers' activities on the platform such as searching for jobs and interviews and calculate a score based on frequency of search or interviews and time since starting their current job. This metric reflects how quickly and earnestly they want to change jobs. We will also need to conduct surveys to let workers rate their job satisfaction, and interviews at the same time to understand their motivation. We can then correlate the two sets of data to see if one is reflective of the other. This hypothesis is important to our work. We plan to work on it as soon as possible.

Banning employers

Another question is whether we should ban certain employers if they are suspected or proven abusive. In the initial rollout of Golden Dreams, Issara screened every employer before they could get on the platform. It is important to have some confidence in the quality of

employers so that we know we are helping workers and mitigating risks when possible. it also helps workers establish trust in the platform.

This may not be a sustainable and scalable solution as the number of employers increases. We will need some automated ways of flagging potentially abusive employers for investigation. This is one of the motivations for developing our joint matching and inference algorithm. It will be especially alarming when there are huge gaps in inferred scores and written reviews and can suggest deception. Thus, this method provides guidance on how human investigation can be prioritized. Investigation results, in turn, can be fed into the inference algorithm to benchmark and test our approach and inform future improvements.

It remains an open question how uncertainty in the inferred score should be incorporate into this process or even decisions to ban certain employers. Because it doesn't involve human subject experimentation, we can potentially use the upper confidence bound to quickly remove high risk employers. It will be up to the expertise of Issara Institute, or other NGOs and international organizations, to decide specific thresholds and standard for removing employers.

Couples

Job seekers can come as a family or couple, and thus prefer to work in the same location or even employers that provide couple housing. In his review of market design literature ([73]), Roth pointed out that it is very difficult to find stable outcomes when there are couples in the market. When redesigning the National Resident Matching Program (NRMP), He and collaborators mitigated this problem by adding a second round of problem resolution to address unstable outcomes one by one. The new process produces stable outcomes most of the time, but the set of possible outcomes can be very small. After observing actual matching platforms of medical residents, he conjectures that as the market grows bigger and if the proportion of couples

88

remain stable, we are more likely to have at least some stable matchings in the end. Therefore, we remain hopeful that even if the problem of couple matching occur in Golden Dreams, we will be able to address it. This solution will not be part of our initial solution due to the complications it introduces both to implementation, and to the understanding and trust of market participants.

Empirical benchmark

We need empirical benchmarks to understand the performance of our solution and make sure it is helpful. It requires measuring key performance indicators (KPIs) before and after the deployment of our solution. These KPIs should capture how easy it is for employers and job seekers to use the platform, satisfaction with employment outcomes in the beginning and in the long-term, accuracy of information we provide, whether the solution is helpful to people's overall goals, etc. The KPIs also need to be monitored and adjusted as the market evolves.

For ease of use, for example, we can (after the initial learning period) measure the overall number of hires versus time spent in recruitment for employers; average time spent on platform until job offer for job seekers; overall time spent on platform per person (longer time probably means people find it useful); completion rate (how many people started using the platform versus how many ended up extending or receiving job offers); etc.

For job satisfaction, we can measure the average turnover rate; time until new job search; frequencies of search or interviews before landing new jobs; the percentage of users who recommend the platform to friends; etc. These are more about initial job satisfaction and can be complemented by surveys and interviews on long term job satisfaction. We can also monitor Issara's Worker Voice platform for reports of job satisfaction and labor exploitation, look at their trends, and see if there is any difference between Golden Dreams users versus those who are not.

We will rely more on interviews and surveys to answer questions about accuracy of information provided and whether the platform helps people's long-term goals. We can also learn about the accuracy of information by looking at Worker Voice records to see whether workers call Issara for help with labor exploitation conditions, despite joining employers with low risk scores. In the long-term, we may see a gradual decline in the ratio of new work visa applications from recruitment agencies (a.k.a MOU requests) versus the number of hires on Golden Dreams, as Golden Dreams increasingly satisfy labor demands from employers more quickly than the MOU process. This number will be highly noisy, though, and impacted by many external factors.

Chapter 10: Discussion

Economics, engineering, and policy

Nobel Prize Laureate Alvin Roth have written several review articles on the design of markets and microeconomics as an engineering field. In 1991, he wrote [74]:

"... the real test of our success will be not merely how well we understand the general principles that govern economic interactions, but how well we can bring this knowledge to bear on practical questions of microeconomic engineering..."

This statement kickstarts a new era where economics, computation, engineering, and policy making are combined to help humanity navigate the information age. With advances in information technology, communication has never been so easy and accessible. Marketplaces for information, transportation, accommodation, work, travel, and food are everywhere and yet they fit into our pockets. Increasingly, their design and performance affect our daily life and fundamental human rights. How does entrepreneurs, academics, and regulators come together to reimagine these fabrics of modern society? What are some good principles to follow in designing and evaluating these marketplaces?

In another review in 2007 ([75]), Roth wrote that the engineering of marketplaces should concern itself with three fundamental problems and one constraint:

1) *Thickness* – are there enough interested participants coming to the market and ready to transact with one another? Otherwise, the market is not useful.

2) *Congestion* – are there too many participants in the market that renders making choices overwhelming? The market would not be effective in that case.

3) *Safety* – the market would be frustrating if the following are not satisfied.

a. Is it safe for participants to be focused on the market instead of considering outside options?

b. Is it safe to be honest and straightforward in the market instead of trying to game the system?

4) *Repugnancy* – are some transactions not morally acceptable by society?

As we have seen in the labor market case, we would like to suggest *uncertainty* as another important problem in marketplace design. It is common participants to have to make decisions when there is uncertainty about one's own preference and about information on the counterparts. It affects the behavior and wellbeing of market participants. Proper design to address the uncertainty underpins both thickness and safety – if information on the market cannot be trusted, some may turn away from the market, reducing its thickness; some may conduct strategic behaviors to gain private information or affect the prevailing belief in the market to gain advantages. Moreover, what are the ethical principles around uncertainty and information disclosure? Should we disclose information with high uncertainty, which may lead to confusion, chaos, and potential harm, or should people's right to know prevail? Traditional markets depend on property rights to function. To answer these new questions, we need clear definitions of the rights of stakeholders in markets.

Market platforms as regulators

Market platforms, as we have seen, play huge roles in society. They have private rules in addition to what is currently written the law and regulations. The rules, in our case, can include when employers are penalized or banned from the platform due to inferred risk or violations; how much should employers know about the calculation behind these penalties to balance their right to know, just treatment, and appeal, versus the risk of gaming the system once details about

the calculation is public; what kind of information should be protected or disclosed and how that affects people's rights to know versus privacy and right not to be discriminated. Platform designers, as a result, becomes private regulators. Their decisions affect the rights and wellbeing of all participants.

We consider a few rights, the rights regarding automated decision-making (and profiling) and to an unbiased treatment, the right to not be unknowingly under to human subject experimentation, and the right to a free and just work environment. We note that the rights to a free and just work environment and not be subject to labor exploitation or human trafficking is a more fundamental right and is the basis of our work. Thus, we focus discussion here on other rights.

Rights regarding automated decision-making and profiling

Article 22 of the European Union General Data Protection Regulation (GDPR) ([76]) states that:

> "1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

unless the decision

    a. "is necessary for entering into, or performance of, a contract between the data subject and a data controller;"

    b. "is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or"

    c. "is based on the data subject's explicit consent."

If such automated decision-making is required, the processer needs to ensure awareness of this processing and the right to appeal ([76], [77]). The UK GDPR *right to be informed* requires that data subjects must know what data is being collected, the purpose of this process, and who it will be shared with ([78]). Thus, employers would need to know that they are being scored for abuse risk based on job seekers' behavior on the platform, and that this information will be shared with job seekers. However, this regulation only applies to personal data about individuals, but not to companies or recruitment agencies, unless the data can be traced to specific employees ([79]). It does not apply to our case of profiling employers for the most part, but it is nonetheless still relevant to our consideration when balancing the rights of different stakeholders.

Rights regarding bias in automated decision-making

The GDPR focuses more on data and less on inference and decisions made based on them. While data protection is gaining momentum, our progress on thinking about bias in the use of data is limited. We applaud the promising progress when at least 17 states introduced bills or resolutions regarding the regulation of artificial intelligence, all but 4 states do have any active bill ([80]). Moreover, only three states, New Jersey, New York, and California, have any mention of discrimination in those bills.

Senator Wyden of Oregon, along with Sen. Cory Booker, D-N.J., and Rep. Yvette D. Clarke, D-N.Y., introduced the Algorithmic Accountability Act in 2019, requiring companies to study and fix the "inaccurate, unfair, biased or discriminatory decisions" in algorithms ([81]). Yet it didn't progress past the committee level either in the House or in the Senate. Most recently, Sen. Edward J. Markey (D-Mass.) and Congresswoman Doris Matsui (CA-06) introduced the Algorithmic Justice and Online Platform Transparency Act of 2021, also with a clause on eliminating bias in algorithmic decision-making ([82]).

An UK government report ([83]) warns that, while the use of algorithms in recruitment can have benefits such as the elimination of human bias, the use of historical data can replicate human bias in the past. It finds that compliance with relevant guidance is mixed, and suggests the governing body, the Equality and Human Rights Commission (EHRC), to update its guidance on the use of algorithms in recruitment.

These considerations are relevant to both employers and job seekers. Employers will be scored for labor exploitation risks. Are these risks biased? Protected traits such as race, gender, disability status, etc. do not apply to companies. However, they would certainly be biased if we present risk scores not based on our best estimates, as we did in chapter 8, but rather on an intentionally biased "upper confidence bound" as in standard MAB literature.

As for job seekers, the issue is both real and prevalent in the use of algorithms in recruitment. They are likely about protected traits such as race, gender, disability standard, and national origin. We must carefully guard against possibilities of bias through algorithmic and human audits and through periodical field research and interviews. More details on this in the recommendation section.

On a more nuanced level, Sandra Wachter and Brent Mittelstadt proposed the *Right to Reasonable Inference* ([84]). They ask the questions of 1) why certain data is a relevant basis to make inference upon; 2) why such inference is required to the purpose of said system; and 3) whether the data and inference methods are accurate and reliable. In our case, we need to know whether the proxies we use to measure worker satisfaction and infer labor exploitation risk, such as job searches and turnover time, are truly reflective of our intended purpose.

In addition to informed consent and several other requirements, the Institutional Review Boards (IRB) review process also require that risks to experimental subjects are "minimized" and "reasonable in relation to anticipated benefits … to subjects and the importance of the knowledge" ([85]). We would certainly require informed consent before any deployment of a matching and inference algorithm that affects the basis on which job seekers make decisions as well as the result from those decisions. Because our inference result is about forced labor, its reliability and accuracy carry inherent risk. We believe that any incentivization of exploration, whether through monetary reward or the control of information disclosed, does not "minimize" risk. It may not even be reasonable in relation to the subjects because the benefits are spread across future job seekers and the risk is assumed by individuals. Thus, we think it is better to avoid this kind of intervention and instead provide our best estimates to people for them to make their own choices.

Policy recommendations

We believe the Golden Dreams platform can be immensely helpful towards mitigating the labor exploitation problem by addressing one of its root causes, labor stress induced by unmet labor needs, as we have shown in part one. Yet, we recognize there are many more root causes of labor abuse and we need more solutions. We will also benefit from better support on the creation, rollout, and maintenance of Golden Dreams. There are a few ways the global labor community can help.

1) Leading international organizations and industry alliances can facilitate development of technology solutions to labor issues by creating innovation programs. Such programs can include consolidation and publication of resources detailing challenges

96

faced by the industry, incubation hubs and accelerators that help innovation teams

getting the ideas off the ground, funding support, and mentorship. These

organizations can also establish internal teams that can help coordinate this effort and

accumulate institutional knowledge. Through such programs, they can offer their

guidance, expertise, and perspectives at various stages of innovation.

2) Brands and their industry alliances can establish internal innovations teams and

corporate venture arms to incubate innovative solution within or outside of

themselves. Their funding should not be limited to venture-based entrepreneurship,

but also social responsibility programs and grants for NGOs and other groups.

3) Brands and their industry alliances can establish standards about such solutions,

including governing principals about such solutions and how suppliers can signify

compliance to labor standard by adopting such solutions. For example, they can

reward their suppliers that use Golden Dreams labor marketplace, which will then

provide employer compliance risk scores, and monitor their own suppliers on the

Worker Voice channel operated by Issara. The reward can be done through

preferential purchase orders. Or, for example, employers who can prove timely and

fair payment of wages can be guaranteed a price above the fair wage payment in

contract negotiation.

Chapter 11: Conclusion

In the first part of this work, we studied the role market frictions in low-skilled labor recruitment play in aggravating human trafficking and forced labor outcomes. We leverage a unique dataset of unmet labor needs in the form of formal migrant worker request and worker reported abuse from a real-time continuously monitored hotline maintained by Issara Institute. Using an instrumental variable approach, we found that labor abuse and trafficking peak when mismatches between local supply and demand for low-skilled labor are unexpectedly large. This result highlights the need for proactive interventions that help suppliers better plan for and adapt to sudden surge in labor demand as a result of purchasing order fluctuations.

As such, we work with Issara Institute on the 2.0 version of Golden Dreams, Issara's recruitment marketplace for Southeast Asia workers. it has already been released and used by workers in Thailand and is gradually being improved and rolled out to more regions. We formulate what the matching process should look like for Golden Dreams, and how we can learn about labor exploitation risks from the process. The goal is to make the platform as easy to use as possible for both workers and employers so that it can help the most people. Through its adoption, we hope to provide more options to workers, help employers meet their labor demand surge, and identify and mitigate labor exploitation risks in the system. The algorithm is guaranteed to return a matching and it has some of the desirable properties in marketplaces.

1) The results are **stable**. No workers or employers can get a different match that also prefers them, so they are safe to accept and stay with their match at least until they learn about things they didn't know before.

2) The results are **optimal** for workers among stable outcomes. This is guaranteed by the DA algorithm.

3) The results are **strategy-proof** (i.e., **safe and straightforward**) for workers and should be almost strategy-proof for employers. This is discussed in the strategic behavior section of chapter 9.

4) The resulting matches are **low regret**. Our conjecture is that we can still achieve $\log(T)$ regret even though we do not incentivize exploration, because the market competition and diversity of job seekers bring natural exploration. We are still actively working on this problem.

There are practical considerations to make sure the labor marketplace works properly and addresses real issues. We discuss the ease of use especially in regard to ranking a potentially overwhelming number of jobs, mitigating undesirable strategic behaviors, accuracy and reliability of our proxies for outcome and our inference, screening and penalty of abusive employers, matching for couples, and lastly the creation of benchmarks.

Finally, we suggest that concerned companies, NGOs, and industry alliances actively participate in the creation and management of this marketplace as well as other and future technical solutions. They can provide funding to expedite the development of such solutions and ensure their scale and proper management and maintenance. They can participate in the management of such solutions by offering their perspectives. They can also incentivize and reward suppliers to use these solutions to solve their business problems, such as recruitment, and to signify their compliance to global standards. Solutions that can help us address labor exploitation issues can include fair labor recruitment platforms, grievance mechanisms that are empowering, helpful, and censor-resistant, and purchasing order planning and management systems that ensure a smoother and low-stress purchase process for brands and suppliers alike.

We caution that such technology solutions will not be sufficient without a broader rethink of the whole system. Brands need to be fairer in the procurement process by providing reasonable and timely payment and enough time and information for the production process. They should invest in cultural change within and across the industry. They can incorporate key performance metrics that are conducive to fair labor conditions in the supply chain. They can establish more consistent information disclosure standard to encourage broader cultural change, and leverage technology solutions to help the entire industry meet the standard. Last by not least, we need to design, build, and maintain these systems and solutions in a way that not only help solve labor exploitation problems through the abovementioned measures, but also secure, privacy-aware, unbiased, and understandable to users.

Bibliography

1. International Labour Organization. "Global Estimates of Modern Slavery: Forced Labour and Forced Marriage." Report, September 19, 2017. http://www.ilo.org/global/publications/books/WCMS_575479/lang--en/index.htm.
2. Muhamad, Tauvik, and Muhammad Nour. "Migrant Fishers Prone to Modern Slavery." Comment, December 20, 2021. http://www.ilo.org/jakarta/info/public/fs/WCMS_832460/lang--en/index.htm.
3. Lind, Dara. "Forced Labor in America: Thousands of Workers Are Being Held against Their Will." Vox, October 22, 2014. https://www.vox.com/2014/10/22/7024483/labor-trafficking-immigrants-exploitation-forced-us-agriculture-domestic-servants-hotel-workers.
4. International Organization for Migration. "Migrants and Their Vulnerability to Human Trafficking, Modern Slavery and Forced Labour - World," July 26, 2019. https://reliefweb.int/report/world/migrants-and-their-vulnerability-human-trafficking-modern-slavery-and-forced-labour.
5. S Chantavanich, S Laodumrongchai, C Stringer, Under the shadow: Forced labor among sea fishers in Thailand. *Mar. Policy* 68, 1–7 (2016).
6. D Tickler, et al., Modern slavery and the race to fish. Nat. Commun.9, 4643(2018).
7. International Labour Organization. "Ending Forced Labour by 2030: A Review of Policies and Programmes." Report, December 10, 2018. http://www.ilo.org/global/topics/forced-labour/publications/WCMS_653986/lang--en/index.htm.
8. Lucille Delahunty. "Strengthening Cross Border Collaboration between Myanmar and Thailand for Children," December 20, 2019. https://www.unicef.org/myanmar/stories/strengthening-cross-border-collaboration-between-myanmar-and-thailand-children.
9. International Labour Organization. "Memorandum of Understanding between the Government of the Kingdom of Thailand and the Government of the Union of Myanmar on Cooperation in the Employment of Workers," June 21, 2003. http://www.ilo.org/asia/info/WCMS_160932/lang--en/index.htm.
10. International Labour Organization. *Recruitment Fees and Related Costs: What Migrant Workers from Cambodia, the Lao People's Democratic Republic, and Myanmar Pay to Work in Thailand*, 2020. https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/documents/publication/wcms_740400.pdf.
11. Human Rights Watch, From the Tiger to the Crocodile (https://www.hrw.org/report/2010/02/23/tiger-crocodile/abuse-migrant-588workers-thailand) (2010).
12. Amnesty International, Nepal: Turning people into profits: Abusive recruitment, trafficking, and forced labor of Nepali migrant workers (2017).
13. Issara Institute, Developing a Financially Viable Ethical Recruitment Model: Prospects for the Myanmar-Thailand Recruitment Channel (2018).
14. International Labor Organization, The Migrant Recruitment Industry: Profitability and unethical business practices in Nepal, Paraguay and Kenya

(http://www.ilo.org/global/topics/fair-recruitment/WCMS_574484/lang–en/index.htm) (2017).

15. Verité, Forced Labor in the Production of Electronic Goods in Malaysia: A Comprehensive Study of Scope and Characteristics (2014).

16. United Nations Office of the High Commissioner for Human Rights(UNOHCHR),OHCHR | Protocol to Prevent, Suppress and Punish Trafficking in Persons (https://www.ohchr.org/en/professionalinterest/pages/protocoltraffickinginpersons.aspx) (2000).

17. Issara Institute. "Top 5 Labour Abuses in 2019-2020- A Field-Based Analysis of Worker Reporting and Business Responses In Thailand," January 2021. https://respect.international/top-5-labour-abuses-in-2019-2020-a-field-based-analysis-of-worker-reporting-and-business-responses-in-thailand/.

18. World Bank. "Better Migration Systems Can Propel Thailand to Grow Faster." *World Bank* (blog), May 16, 2018. https://www.worldbank.org/en/news/press-release/2018/05/16/better-migration-systems-can-propel-thailand-to-grow-faster.

19. Walk Free Foundation. 2019. Global Slavery Index. Online. URL https://www.globalslaveryindex.org/ 2018/data/maps/#prevalence. [Last accessed January 7, 2022].

20. Human Rights Watch. "'Paying for a Bus Ticket and Expecting to Fly': How Apparel Brand Purchasing Practices Drive Labor Abuses." Human Rights Watch, April 23, 2019. https://www.hrw.org/report/2019/04/23/paying-bus-ticket-and-expecting-fly/how-apparel-brand-purchasing-practices-drive.

21. Caro, Felipe, Leonard Lane, and Anna Sáez de Tejada Cuenca. "Can Brands Claim Ignorance? Unauthorized Subcontracting in Apparel Supply Chains." *Management Science* 67, no. 4 (April 1, 2021): 2010–28. https://doi.org/10.1287/mnsc.2020.3679.

22. Locke, Richard M. *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy*. Cambridge Studies in Comparative Politics. Cambridge: Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139381840.

23. G Distelhorst, J Hainmueller, RM Locke, Does Lean Improve Labor Standards? Management and Social Performance in the Nike Supply Chain. *Manag. Sci.* 63, 707–728 (2017).

24. Lee, Hau, and Sheila Melvin. "Everything Is Connected: A New Era of Sustainability at Li & Fung." Stanford Graduate School of Business, 2015. https://www.gsb.stanford.edu/faculty-research/case-studies/everything-connected-new-era-sustainability-li-fung.

25. "Migrant Worker Shortage Threatens Key Thai Exports, Economic Recovery." *Reuters*, July 30, 2021, sec. Bonds News. https://www.reuters.com/article/thailand-economy-migrants-idUKL4N2OV0R3.

26. "Thailand Opening Doors to More Foreign Workers to Revive Economy." *Bloomberg.Com*, November 17, 2021. https://www.bloomberg.com/news/articles/2021-11-17/thailand-opening-doors-to-more-foreign-workers-to-revive-economy.

27. Centro de los Derechos del Migrante, Inc. "Ripe for Reform: Abuse of Agricultural Workers in the H-2A Visa Program." Centro de los Derechos del Migrante, Inc., 2020. https://cdmigrante.org/wp-content/uploads/2020/04/Ripe-for-Reform.pdf.

28. Human Rights Watch. "Global Treaty to Protect Forced Labor Victims Adopted." *Human Rights Watch* (blog), June 11, 2014. https://www.hrw.org/news/2014/06/11/global-treaty-protect-forced-labor-victims-adopted.

29. International Labour Organization. "Protocol P029 - Protocol of 2014 to the Forced Labour Convention, 1930," 2014. https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:P029.

30. "Ending Forced Labour by 2030: A Review of Policies and Programmes." Report, December 10, 2018. http://www.ilo.org/global/topics/forced-labour/publications/WCMS_653986/lang--en/index.htm.

31. Congressional Research Service. "Section 307 and Imports Produced by Forced Labor." Congressional Research Service, December 27, 2021. https://crsreports.congress.gov/product/pdf/IF/IF11360.

32. "EU to Propose Ban on Products Made by Forced Labour - von Der Leyen." *Reuters*, September 15, 2021, sec. Commodities News. https://www.reuters.com/article/eu-commission-labour-idAFS8N2O6018.

33. Apple, Inc. "People and Environment in Our Supply Chain - 2021 Annual Progress Report." Apple. Accessed September 27, 2021. https://www.apple.com/supplier-responsibility/pdf/Apple_SR_2021_Progress_Report.pdf

34. "How a Deeper Dive into Our Supply Chain Led to a New Migrant Worker Standard - Patagonia," June 3, 2015. https://www.patagonia.com/stories/the-unacceptably-high-cost-of-labor-a-new-migrant-worker-standard-from-patagonia/story-17743.html.

35. Unilever. "Human Rights in Our Extended Value Chain." Accessed September 28, 2021. https://www.unilever.com/planet-and-society/respect-human-rights/human-rights-in-our-extended-value-chain/.

36. CIERTO. "About | Cierto." Accessed November 16, 2021. https://www.ciertoglobal.org/about/.

37. Verité. "Recruitment and Hiring Practices in the Philippine Tuna Handline Fishing Sector." Verité, September 2020. https://s38965.pcdn.co/wp-content/uploads/2020/09/Verite_Recruitment-and-Hiring-Practices-in-the-Philippine-Tuna-Handline-Fishing-Sector.pdf.

38. Polaris Project. "2019 Data Report: The U.S. National Human Trafficking Hotline." Polaris Project, 2019. https://polarisproject.org/wp-content/uploads/2019/09/Polaris-2019-US-National-Human-Trafficking-Hotline-Data-Report.pdf.

39. Unseen. "Modern Slavery & Exploitation Helpline Annual Assessment." Unseen, 2019. https://www.unseenuk.org/helpline-statistics/annual-assessment-2019/.

40. The European Union and International Labour Organization. "Global Action to Improve the Recruitment Framework of Labour Migration (REFRAME)." Project, June 2, 2020. http://www.ilo.org/islamabad/whatwedo/projects/WCMS_746771/lang--en/index.htm.

41. Office of the United States Trade Representative. "Thailand." Accessed November 15, 2021. http://ustr.gov/countries-regions/southeast-asia-pacific/thailand.

42. World Bank. "Thailand Product Exports to China 2018 | WITS Data." Accessed November 15, 2021. https://wits.worldbank.org/CountryProfile/en/Country/THA/Year/2018/TradeFlow/Export/Partner/CHN/Product/all-groups.

43. United Nations Office of the High Commissioner for Human Rights (UNOHCHR), Guiding Principles for Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework | UN Global Compact (https://www.unglobalcompact.org/library/2) (2011).

44. MSI Integrity, Not Fit-for-Purpose (2020).

45. Re:Structure Lab, Due Diligence And Transparency Legislation, (Sheffield, Stanford, and Yale Universities, Sheffield), Technical report (2021).

46. Issara Institute. "Golden Dreams." *Golden Dreams Fact Sheet* (blog). Accessed December 13, 2021. https://www.issarainstitute.org/golden-dreams.

47. Issara Institute. *Golden Dreams Recruitment Marketplace - Overview*, 2021. https://www.youtube.com/watch?v=yrROikJl91w.

48. Hautala, Laura. "Amazon's Never-Ending Fake Reviews Problem, Explained." CNET. November 23, 2021. https://www.cnet.com/features/amazons-never-ending-fake-reviews-problem-explained/.

49. D'Onfro, Jillian. "A Whopping 20% Of Yelp Reviews Are Fake." Business Insider. September 25, 2013. https://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9.

50. Oremus, Will, Chris Alcantara, Jeremy B. Merrill, and Artur Galocha. "How Facebook Shapes Your Feed." *Washington Post*, October 26, 2021. https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/.

51. McMahan, H. Brendan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, et al. "Ad Click Prediction: A View from the Trenches." In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1222–30. Chicago Illinois USA: ACM, 2013. https://doi.org/10.1145/2487575.2488200.

52. Pathak, Deepak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. "Curiosity-Driven Exploration by Self-Supervised Prediction." *ArXiv:1705.05363 [Cs, Stat]*, May 15, 2017. http://arxiv.org/abs/1705.05363.

53. Hayek, F. A. (1945). The use of knowledge in society. *The American economic review, 35*(4), 519-530.

54. Beam, Emily A. "Do Job Fairs Matter? Experimental Evidence on the Impact of Job-Fair Attendance." *Journal of Development Economics* 120 (May 1, 2016): 32–40. https://doi.org/10.1016/j.jdeveco.2015.11.004.

55. Dahyeon Jeong. "Using SMS to Transform Agricultural Labor Markets in Tanzania," December 18, 2019. https://blogs.worldbank.org/impactevaluations/using-sms-transform-agricultural-labor-markets-tanzania-guest-post-dahyeon-jeong.

56. Gale, D., and L. S. Shapley. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly* 69, no. 1 (January 1, 1962): 9–15. https://doi.org/10.1080/00029890.1962.11989827.

57. Roth, Alvin. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy* 92, no. 6 (1984): 991–1016.

58. Abdulkadiroglu, Atila, Parag Pathak, Alvin E. Roth, and Tayfun Sonmez. "Changing the Boston School Choice Mechanism." Working Paper. Working Paper Series. National Bureau of Economic Research, January 2006. https://doi.org/10.3386/w11965.

59. Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica* 85, no. 5 (2017): 1373–1432. https://doi.org/10.3982/ECTA13925.

60. Liu, Lydia T., Horia Mania, and Michael I. Jordan. "Competing Bandits in Matching Markets." *ArXiv:1906.05363 [Cs, Stat]*, July 12, 2020. http://arxiv.org/abs/1906.05363.

61. Slivkins, Aleksandrs. "Introduction to Multi-Armed Bandits." *ArXiv:1904.07272 [Cs, Stat]*, September 29, 2019. http://arxiv.org/abs/1904.07272.

62. Frazier, Peter, David Kempe, Jon Kleinberg, and Robert Kleinberg. "Incentivizing Exploration." In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, 5–22. Palo Alto California USA: ACM, 2014. https://doi.org/10.1145/2600057.2602897.

63. Chen, Bangrui, Peter Frazier, and David Kempe. "Incentivizing Exploration by Heterogeneous Users." In *Proceedings of the 31st Conference On Learning Theory*, 798–818. PMLR, 2018. https://proceedings.mlr.press/v75/chen18a.html.

64. Mansour, Yishay, Aleksandrs Slivkins, and Vasilis Syrgkanis. "Bayesian Incentive-Compatible Bandit Exploration." *ArXiv:1502.04147 [Cs]*, May 2, 2019. http://arxiv.org/abs/1502.04147.

65. Cen, Sarah H., and Devavrat Shah. "Regret, Stability, and Fairness in Matching Markets with Bandit Learners." *ArXiv:2102.06246 [Cs, Stat]*, February 11, 2021. http://arxiv.org/abs/2102.06246.

66. Bastani, Hamsa, Mohsen Bayati, and Khashayar Khosravi. "Mostly Exploration-Free Algorithms for Contextual Bandits." *ArXiv:1704.09011 [Cs, Stat]*, April 18, 2020. http://arxiv.org/abs/1704.09011.

67. Roth, Alvin E. "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." *Econometrica* 54, no. 2 (1986): 425–27. https://doi.org/10.2307/1913160.

68. Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70, no. 1 (1983): 41–55. https://doi.org/10.2307/2335942.

69. Roth, Alvin E. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research* 7, no. 4 (November 1982): 617–28. https://doi.org/10.1287/moor.7.4.617.

70. Roth, Alvin E., and Elliott Peranson. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review* 89, no. 4 (September 1999): 748–80. https://doi.org/10.1257/aer.89.4.748.

71. Ashlagi, Itai, and Yannai A. Gonczarowski. "Stable Matching Mechanisms Are Not Obviously Strategy-Proof." *Journal of Economic Theory* 177 (September 2018): 405–25. https://doi.org/10.1016/j.jet.2018.07.001.

72. Klemperer, Paul. "What Really Matters in Auction Design." *Journal of Economic Perspectives* 16, no. 1 (March 2002): 169–89. https://doi.org/10.1257/0895330027166.

73. Roth, Alvin E. "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions." Working Paper. Working Paper Series. National Bureau of Economic Research, July 2007. https://doi.org/10.3386/w13225.

74. Roth, Alvin E. "Game Theory as a Part of Empirical Economics." *The Economic Journal* 101, no. 404 (1991): 107–14. https://doi.org/10.2307/2233845.

75. Roth, Alvin E. "What Have We Learned From Market Design?" Working Paper. Working Paper Series. National Bureau of Economic Research, October 2007. https://doi.org/10.3386/w13530.

76. EU. "Art. 22 GDPR – Automated Individual Decision-Making, Including Profiling." *General Data Protection Regulation (GDPR)* (blog), April 27, 2016. https://gdpr-info.eu/art-22-gdpr/.

77. UK Information Commissioner's Office. "Rights Related to Automated Decision Making Including Profiling." ICO, November 3, 2021. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/.

78. UK Information Commissioner's Office. "Right to Be Informed," August 2, 2018. https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-be-informed-1-0.pdf.

79. European Commission. "Do the Data Protection Rules Apply to Data about a Company?" Text. European Commission - European Commission. Accessed January 10, 2022. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/do-data-protection-rules-apply-data-about-company_en.

80. National Conference of State Legislatures. "Legislation Related to Artificial Intelligence," January 5, 2022. https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx.

81. U.S. Senator Cory Booker of New Jersey. "Booker, Wyden, Clarke Introduce Bill Requiring Companies To Target Bias In Corporate Algorithms," April 10, 2019. https://www.booker.senate.gov/news/press/booker-wyden-clarke-introduce-bill-requiring-companies-to-target-bias-in-corporate-algorithms.

82. U.S. Senator Ed Markey of Massachusetts. "Senator Markey, Rep. Matsui Introduce Legislation to Combat Harmful Algorithms and Create New Online Transparency Regime," May 27, 2021. https://www.markey.senate.gov/news/press-releases/senator-markey-rep-matsui-introduce-legislation-to-combat-harmful-algorithms-and-create-new-online-transparency-regime.

83. Centre for Data Ethics and Innovation. "Review into Bias in Algorithmic Decision-Making." Centre for Data Ethics and Innovation, November 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf.

84. Sandra Wachter and Brent Mittelstadt. "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI." *Oxford Law Faculty* (blog), October 9, 2018. https://www.law.ox.ac.uk/business-law-blog/blog/2018/10/right-reasonable-inferences-re-thinking-data-protection-law-age-big.

85. Human Research Protection Program. "Guidance: Requirements for IRB Review and Approval." University of California, Los Angeles, April 14, 2021. https://ora.research.ucla.edu/OHRPP/Documents/Policy/12/IRB_Review_Requirements.pdf.