cesifo WORKING PAPERS

6922
2018

February 2018

The International Organization of Production in the Regulatory Void

Philipp Herkenhoff, Sebastian Krautheim

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6922*
Category 8: Trade Policy

The International Organization of Production in the Regulatory Void

Abstract

Over the last decades, the internationalization of the value chain has allowed firms to exploit cross-country differences in environmental and labor regulation (and enforcement) in ways that have led to a large number of NGO campaigns and consumer boycotts criticizing ‘unethical’ practices. How do potential ‘unethical’ cost savings on the one hand and the threat to reputation and sales on the other interact with the international organization of production? In this paper we introduce North-South differences in regulation, a cost-saving ‘unethical’ technology and consumer boycotts into a standard property rights model of international production. Contracts are incomplete, so that a firm has limited control over both investments and (un)ethical technology choices of both foreign affiliates and suppliers along the value chain. We show that international outsourcing and ‘unethical’ production are linked through a novel unethical outsourcing incentive, for which we also provide empirical support: a high cost advantage of ‘unethical’ production in an industry and a low regulatory stringency in the supplier's country favor international outsourcing (as opposed to vertical FDI). We also provide a microfounded model of investment and pricing under incomplete contracts when the production technology is a credence characteristic of the final good and an NGO investigates firms and may initiate a consumer boycott.

JEL-Codes: D210, D230, F120, F230, F610, J810, L220, L230, L310, L500, Q530.

Keywords: multinational firms, international outsourcing, property rights theory of the firm, ethical production, labor standards, pollution, consumer boycotts, credence goods, NGOs.

Philipp Herkenhoff
LMU Munich / Germany
Philipp.Herkenhoff@econ.lmu.de

*Sebastian Krautheim**
University of Passau / Germany
sebastian.krautheim@eui.eu

*corresponding author

February 2018
We would like to thank Gani Aldashev, Bård Harstad, Thierry Verdier, participants of the 19th International Economics Workshop in Göttingen, the ‘Non-Profits, Governments, and Organizations Workshop’ at the Graduate Institute, Geneva, as well as seminar participants at the University of Hohenheim, LMU Munich, Paris School of Economics, the University of Trier and the Université Catholique de Louvain for fruitful discussions and comments. The Online Appendix is available upon request from the authors.

1 Introduction

The past three decades were characterized by an unprecedented fragmentation and geographical dispersion of production. Value chains span all over the globe and even firms with a strong national branding have highly segmented international supply chains. The fact that low trade and information costs allow firms to exploit cross-country differences in factor prices is well established in the economics literature. But this profound change in the locus and organization of production also allows firms to exploit differences in environmental and labor regulation as well as enforcement capacity across countries. With national regulatory regimes and multinational production, these firms operate in a regulatory void (Short, 2013).

In this context, multinational firms and their suppliers are frequently accused of cutting costs at the expense of the environment, local workers and future generations. In a large number of cases such allegations of 'unethical' or 'immoral' practices have led to NGO campaigns and consumer boycotts inflicting sizable damage to reputation, sales and value of the targeted firms.[1]

Many global industry leaders have faced consumer boycotts (or the threat thereof) due to issues along their international supply chains. Examples include Nike for sweatshops in Indonesia (Harrison and Scorse, 2010), Apple and Samsung for abusive work conditions and environmental pollution in their supplier factories in China (China Labor Watch, 2018, Bloomberg, 2018, and China Labor Watch, 2012), McDonalds, Pepsico, Nestlé, Unilever and Procter and Gamble for rainforest destruction by their palm oil suppliers in Indonesia (Rainforest Action Network, 2017, Neslen, 2017), Coca-Cola for child labor at sugar cane suppliers in El Salvador (Human Rights Watch, 2004, Smedley, 2014) and toy producers Hasbro and Mattel for labor abuse by their suppliers in China (China Labor Watch, 2015, Groden, 2015). The top five apparel brands Nike, Zara, H&M, Adidas, and Uniqlo[2] have - among many others - agreed to remove hazardous chemicals from their entire supply chain by 2020 following the Detox campaign by Greenpeace (2016). The Clean Clothes Campaign (2017) has documented abusive work conditions at Indian supplier factories of Abercrombie&Fitch, Benetton, C&A, Columbia, Decathlon, Old Navy, Banana Republic, H&M, Levi's, Marks&Spencer, Hilfiger and Calvin Klein. These are just a few examples to illustrate that leading firms in a diverse range of industries are subject to campaigns addressing labor and environmental issues along their international supply chains. Baron (2012) and Krautheim and Verdier (2016) provide additional examples.

Even the most casual observation of these examples reveals that most campaigns criticize the actions of independent suppliers rather than subsidiaries of multinational firms (although both cases exist). So, does the choice between 'ethical' and 'unethical' production interact with the international organization of production? Are independent suppliers more likely to implement 'unethical' technologies than affiliates of multinational firms? From within the political sciences literature on international production and labor rights, (Mosley, 2011, p.7) argues that this is indeed the case: *"MNC-owned global production affects labor rights in a positive fashion, whereas subcontracted production is associated with less respect for worker's rights."* She provides some cross-country evidence consistent with her claim.

Within the field of economics, the literature on the international organization of production based on

[1]King and Soule (2007) and Flammer (2013) document resulting losses in stock market valuation. Suffering demand losses and oversupply due to the sweatshop campaign in May 1998, Spar and La Mure (2003) quote Nike CEO Phil Knight as saying: "The Nike product has become synonymous with slave wages, forced overtime, and arbitrary abuse. I truly believe the American consumer doesn't want to buy products made under abusive conditions."

[2]We take the ranking from KantarMillwardBrown (2017).

the seminal contribution by Antràs (2003) appears to be the natural framework to analyze this relationship. It applies the property rights theory of the firm by Grossman and Hart (1986) and Hart and Moore (1990) to international value chains in a North-South context. Several determinants of the choice between vertical integration and international outsourcing have been analyzed both theoretically and empirically, e.g. capital/headquarter intensity in Antràs (2003), firm heterogeneity in Antràs and Helpman (2004), contractibility of inputs in Antràs and Helpman (2008), routineness in Costinot, Oldenski, and Rauch (2011), and downstreamness of the supplier in the value chain in Antràs and Chor (2013).

In this paper we introduce North-South differences in regulation, a cost-saving 'unethical' technology and consumer boycotts into this literature. We seek to better understand how the boundaries of the firm respond when the implementation of possibly legal but supposedly 'unethical' practices in one country can have repercussions on sales and profits in another country. We find that there is indeed an additional outsourcing incentive that is absent in the previous literature: the cost savings of 'unethical' production alter optimal investments along the value chain and thereby make outsourcing more attractive. This effect is strongest in supplier-intensive sectors and implies that sectors with high potential cost savings of 'unethical' production are more prone to keeping their suppliers at arm's length. We also provide evidence for the empirical relevance of this effect.

We place our analysis in a context where the internationalization of production lets firms locate parts of their value chains in a jurisdiction (the 'Global South') with a more lenient regulation and/or lower enforcement capacity. This allows firms to implement a technology, forbidden in the Global North, which saves costs, but generates an externality on a third party (e.g. local pollution, unsustainable extraction of renewable resources or poor labor, safety and health standards). The first premise of our analysis is that these externalities raise ethical concerns on the side of consumers in the Global North potentially resulting in a consumer boycott of the final product.[3] As the production technology cannot be inferred from the final product, and is difficult or impossible to be verified by final consumers, it constitutes a credence attribute of the final product (Feddersen and Gilligan, 2001 and Baron, 2011). In the absence of international regulation addressing the market failures associated to credence goods, social activists can respond to this international governance deficit (Gereffi and Mayer, 2006) by initiating consumer boycotts to influence the production technology of firms along the value chain.[4]

The second premise of our analysis is that production along international value chains is characterized by incomplete contracts. This is the central assumption of the property rights theory of international production suggested by Antràs (2003) and appears very natural for production in a North-South context.

[3]We do not take any normative stand on what 'ethical' or 'unethical' practices are. In our analysis, the defining feature of an 'unethical' technology is simply that it saves costs but may trigger a consumer boycott. There is ample empirical evidence both from surveys (O'Rourke, 2005 and Loureiro and Lotade, 2005) and from field experiments with real purchasing decisions (e.g. Hiscox and Smyth, 2011 and Hainmueller and Hiscox, 2012) that consumers do care about such issues and have a higher willingness to pay for ethical products. Moreover, Basu and Tzannatos (2003) and Cone (2013) provide evidence that this awareness has increased over the last decades.

[4]While most evidence on NGO campaigns rests on case studies, very recently, more systematic evidence is provided by Hatte and Koenig (2018). For a period from 2010–2014, their raw data contain campaigns of 2949 activists, campaigning against a total of 6893 firms headquartered in 130 countries. Using data on the location of the firm's headquarter, the NGO's headquarter as well as the country in which the criticized action takes place, they show that the international fragmentation of production is also reflected in the activity of advocacy NGO campaigns. On the one hand, they find a strong international dimension of this activity. On the other hand, their triadic gravity analysis shows a strong bias for NGOs to campaign against domestic firms. This pattern is consistent with our modeling approach where a firm from the Global North is confronted by an NGO from the Global North about an action taking place in the Global South.

Issues concerning dispute settlement, place of jurisdiction as well as questions of enforcement across borders arise in this context. Moreover, relationship-specificity of investments along the value chain aggravates these problems as it adds dimensions to the product that are hard to specify ex-ante and difficult to verify by a third party ex-post. In our context, this contractual incompleteness naturally extends to the implementation of technology: no contract effectively binds the supplier to implement the ethical or unethical technology type. The massive difficulties of internationally active firms trying to implement codes of conduct for their suppliers largely backs this assumption.[5]

One central result of our analysis is that the headquarter and supplier intensities of the production technology are key determinants both for the organization of production *and* for the choice between the unethical and the ethical technology. We find that, just like in Antràs (2003), a high supplier intensity favors international outsourcing over vertical integration. In addition, we find that a high supplier intensity favors unethical production. This implies that in sectors where the supplier provides an important contribution to the production process one should observe both more outsourcing *and* more unethical production.[6] In line with the hypothesis by Mosley (2011) our model therefore implies an association between international outsourcing and unethical production.

Further analysis of the mechanics of our model reveals a more sophisticated relation between the two. In our model, outsourcing is an instrument for the headquarter to alleviate the underinvestment of the supplier. Unethical production increases the gap between the optimal and the actual investment. This aggravated underinvestment under unethical production magnifies the incentive of the headquarter to choose outsourcing compared to the ethical (i.e. the Antràs, 2003) case. We label this the *unethical outsourcing incentive*.

This generates a range of factor intensities for which outsourcing is only chosen because the headquarter anticipates unethical production by the supplier. The supplier's option to implement the unethical technology therefore biases the organizational decision of the firm towards outsourcing. This effect is more pronounced for sectors with stronger incentives for unethical production. Our model therefore implies that sectoral variation in the incentives for unethical production - in terms of model parameters: high unethical cost advantage and a low probability of a boycott - is associated with sectoral variation in the organizational form *after* controlling for factor intensities. In Section 4.6 we provide empirical evidence for this conditional correlation: controlling for all standard proxies of headquarter intensity, we find that high cost savings from unethical production in an industry are associated with more outsourcing relative to vertical integration.

The unethical outsourcing incentive also implies an interesting tension between aspirations and reality when it comes to the headquarter's actions. In the public debate firms are frequently accused of 'green washing', i.e. claiming to be in favor of ethical production but acting differently. We find that the combination of actually *wishing* to source ethically but *expanding* unethical production can be an equilibrium outcome. This is the case when the headquarter would *like* to oblige the supplier to implement the ethical technology (which it cannot) but *anticipates* unethical production. The headquarter then has an incentive

[5]Nike is a well documented case in point (e.g. Locke, Qin, and Brause, 2007). Other research documenting difficulties of implementing codes of conduct with independent suppliers includes Egels-Zandén (2007), Ruwanpura and Wrigley (2011), and Bird, Short, and Toffel (2017).

[6]Despite this link, our model is rich enough to also feature *ethical outsourcing* and *unethical integration* as equilibrium outcomes.

to maximize cost savings from unethical production, which is achieved by keeping the supplier at arm's length and thereby scaling up unethical production.

With consumers willing to boycott certain products on ethical grounds, information about the implemented production technology is crucial. We assume that technology is a credence attribute of a product - it cannot be inferred from the final product even after consumption. That said, observable firm choices (like investments, quantities and prices) may nevertheless contain information about the underlying technology. A deviation from those investments, quantities or prices that are optimal under ethical production may then indicate that the unethical technology is implemented. In the baseline model we simply impose that any deviation from these observables is interpreted as proof of unethical production and directly triggers a consumer boycott, leading firms to set observables like under ethical production. This allows us to focus on the implications of our model for the international organization of production in the baseline model in Section 2. In Section 3 we provide a microfoundation in which we introduce an activist NGO screening firms for signs of unethical behavior and organizing consumer boycotts in response. We find that - as in the baseline model - in equilibrium unethical firms hide their type by pooling with ethical firms. We show that both the microfounded and the baseline model yield qualitatively identical results.[7]

The need to pool with ethical firms implies that in equilibrium unethical firms set the same investment levels as ethical firms. Given that an unethical supplier faces lower variable costs, the deviation of the *actual* from the *optimal* investment increases. As outsourcing is the only instrument to mitigate the underinvestment by the supplier, outsourcing becomes more attractive under unethical production, generating the unethical outsourcing incentive.

While the (un)ethical technology choice of the supplier depends on the factor intensity of production (and, quite intuitively, the cost advantage of unethical production and the risk of a boycott), it turns out to be *independent* of the organization of production. The reason for this is that the organization of production has two countervailing effects on the attractiveness of unethical production for the supplier, which exactly offset one another. On the one hand, outsourcing scales up cost savings from unethical production through increased investments, on the other, it makes the supplier more vulnerable to a boycott by increasing its share of final revenues. This implies that the headquarter has no instrument to affect the supplier's technology choice. Our model therefore remains very close to the setting in Antràs (2003) and allows us to focus on the prevalent question in this literature: how can one instrument (organization of production) be used to affect one variable (investment incentives) under incomplete contracts? Our setting allows us to analyze how the potentially unethical technology choice of the supplier distorts the use of the instrument by the headquarter, linking unethical production to outsourcing in equilibrium.[8]

As outlined above, our model predicts a link between the incentives for unethical production and outsourcing. To support this prediction empirically, we follow the established literature, e.g. Nunn and Trefler (2013), Antràs and Chor (2013) and Antràs and Yeaple (2014), in using U.S. Census Bureau data

[7] Krautheim and Verdier (2016) analyze endogenous NGO emergence in response to offshoring when technology is a credence attribute. They analyze the signaling game between a potentially unethical firm and consumers where prices and quantities produced can be signals of the technology implemented. A key difference in our setting is that the 'signal' is not the result of a purposeful choice of a single agent (potentially used to signal its type) but the result of a non-cooperative investment of the headquarter and the supplier.

[8] One could consider a second instrument, like an investment into corporate social responsibility (CSR) directly affecting the supplier's ability or incentive to implement the unethical technology. We leave this extension of the model for future research.

on intrafirm trade. We use the standard measure of vertical integration at the product level: the share of U.S. intrafirm imports in total U.S. imports for the years $2007 - 2014$. We correlate this variable with the cost advantage of unethical production and analyze if this relation differs systematically across levels of regulatory stringency of the exporting country.

We suggest a measure of cost savings at the expense of the environment, for which we draw on data from the Annual Survey of Manufactures provided by the U.S. Census Bureau. Since 2007 the survey has recorded the industry-level expenditure on water, sewer, refuse removal, and other non-electric utility payments including the cost of hazardous waste removal. These expenditures are highly sensitive to regulation. We use their share in total costs to proxy for potential cost savings from operating under more lenient regulation (with the US as the benchmark). To measure the level of regulation (and enforcement) in different countries, we use the Environmental Policy Stringency Index computed by the OECD for 33 countries for the years 2007 to 2012, including the six non-member countries Brazil, China, India, Indonesia, Russia, and South Africa. The index combines information on 14 environmental policy instruments that are mainly related to air and climate pollution and is suitable for comparisons across countries.

As predicted by our model, we find a statistically and economically significant negative relationship between our measure of the unethical cost advantage and the share of intrafirm imports in an industry. The relationship is stronger in countries with more lenient regulation. These findings are robust to the inclusion of country-year fixed effects as well as a large number of control variables that the previous literature has found - and our theory predicts - to affect the prevalence of vertical integration.

Our paper contributes to the large literature on the international organization of production pioneered by Antràs (2003). Some of the major contributions are highlighted above and a recent survey is provided by Antràs and Yeaple (2014). Differences in regulation and institutions are not alien to this literature. Antràs and Helpman (2004) assume that integration improves the outside option of the headquarter more in the North than in the South. In Antràs and Helpman (2008), the share of contractible inputs may differ between production locations. In contrast to those papers, we do not consider institutions like rule of law and the protection of property rights, but instead consider differences in environmental regulation and labor rights and their enforcement.

By introducing consumer boycotts and an advocacy NGO into a model of the international organization of production, our paper contributes to bridging the gap between the international economics literature and the literature on private politics started by Baron (2001, 2003). The latter focuses on activists attempting to affect firm behavior not through lobbying for regulation (*public politics*) but through campaigns and boycotts of firms (*private politics*). This literature takes an industrial organization perspective and analyzes the interaction of activists, firms and possibly a regulator under different market structures and allowing for strategic interactions between all parties.[9]

Most related to our approach, some papers have introduced elements of private politics into international economics. Aldashev and Verdier (2009) analyze the international competition for funds among development-oriented NGOs. Aldashev, Limardi, and Verdier (2015) consider the impact of NGO

[9] Some of the main contributions include Innes (2006), Baron and Diermeier (2007), Lyon and Salant (2013), Baron (2010), as well as Baron (2016), and Egorov and Harstad (2017). Closely related to the private politics literature, but with a different focus, are works on the private provision of public goods and corporate social responsibility (CSR) surveyed by Kitzmuller and Shimshack (2012).

campaigns on industry structure in a setting with endogenous mark-ups and monopolistic competition. Krautheim and Verdier (2016) analyze the endogenous emergence of a consumer-financed NGO in response to the offshoring decision of a firm. Kitzmuller (2012) takes the model of Besley and Ghatak (2007), who explicitly model an NGO as a potential provider of a public good, to the international level.

Issues related to private regulation, social activism and NGO-firm interactions in global value chains have received much more attention in political sciences and management studies. This literature finds that social activism is instrumental in the establishment of codes of conduct in multinational supply chains and analyzes further determinants of their success in case studies and more recently in large firm-level datasets.[10]

Our work has some relation to several strands of the international trade literature. First and foremost, Copeland and Taylor (1994) formalize the idea that differences in environmental regulation affect the international location of production. This triggered a large literature on trade (FDI) and the environment which is surveyed in Copeland and Taylor (2004).[11] Our approach has in common with this literature that we view regulatory differences as a driving force of the internationalization of production. This literature, however, does not analyze the international organization of production and, importantly, ignores the feedback effects the implementation of unethical technologies can have on demand when consumer boycotts are possible.[12]

The remainder of this paper is structured as follows. In Section 2, we present our baseline model of unethical production and consumer boycotts and analyze the optimal international organization of production. In Section 3, we analyze an extension of the model featuring advocacy NGOs and asymmetric information, thereby microfounding the relation between consumer boycotts and observables like the organization of production, investments, quantities, and prices. We describe the empirical specification along with the data sources and the results of our empirical analysis in Section 4. Section 5 concludes.

2 A Model of (Un)ethical Sourcing with Incomplete Contracts

In this section, we outline a property-rights model of the boundaries of the firm in the context of international differences in labor or environmental regulation and the risk of consumer boycotts. To facilitate the comparison to the existing literature, we closely follow Antràs (2003) in our baseline setting.[13] Similar to

[10] See e.g. Locke, Kochan, Romis, and Qin (2007), Distelhorst, Hainmueller, and Locke (2017), Ouellet, Short, and Toffel (2015), Distelhorst and Locke (forthcoming) and references therein.

[11] See Aichele and Felbermayr (2015) and references therein for more recent contributions.

[12] We identified two other strands of the literature that resonate with some dimension of our analysis. First, the 'protection for sale' literature based on Grossman and Helpman (1994), which considers the influence of a special interest group on trade policy outcomes. The focus is therefore on *public politics* rather than on *private politics* and on *trade policy* rather than on the *international organization of production*. Second, Eckel and Egger (2009) study the role of trade unions for international investment and production decisions of firms. There are several important differences between advocacy NGOs and trade unions. The former affect firms through demand, tend to be indifferent to survival of the firm and address externalities that usually concern third parties. The latter in turn affect firms on the cost side, vitally depend on firm survival and maximize the utility of their (nationally segmented) members.

[13] We do not include firm heterogeneity like in Antràs and Helpman (2004) in our model, but rather take the original model in Antràs (2003) as a reference point. The reason is that to our knowledge there are no stylized facts concerning correlations of firm size, productivity or quality to the implementation of unethical production, that could guide our modeling. We argue that our mechanism is equally general as the mechanism in Antràs (2003). It should therefore carry over to *any* extension of the original Antràs model, but possibly at the expense of tractability. The attentive reader of our microfoundation spelled out in Section 3 might think that in a model with heterogeneous firms the link between investment choices and boycotts may break down. We argue in Footnote 26 that this is not the case in particular if heterogeneity is modeled as differences

Antràs and Chor (2013), we focus on the analysis of the organizational choice of the headquarter-supplier pair and abstract from an analysis of the industry equilibrium.

2.1 Baseline Model

In this section we outline our baseline model that treats the link between firm choices (like investments, quantities and prices) and the formation of consumers' expectations on the - unobservable - technology implemented in a very stylized way. This allows us to derive our main results on the interaction of consumer boycotts with the international boundaries of the firm. In Section 3 we provide a micro foundation featuring an NGO investigating firms and initiating boycotts. We show that our results from the baseline model hold qualitatively.

2.1.1 Preferences, Consumer Boycotts and Demand

All consumers are located in the Global North. Their preferences are summarized by the following CES aggregate over a large number of symmetric varieties indexed by ω,

$$U = \left(\int_{\omega \in \Omega} y(\omega)^{\alpha} I(\omega) d\omega \right)^{\frac{1}{\alpha}}, \tag{1}$$

with $\alpha \in (0,1)$, Ω being the set of available varieties and $y(\omega)$ representing the quantity consumed of variety ω. These preferences are standard with the exception of the indicator variable $I(\omega)$. It reflects the fact that a firm (and its variety) can be hit by a consumer boycott. In this case the indicator variable takes a value of zero implying that the representative consumer does not derive any utility from its consumption.

Consumers maximize their utility subject to the budget constraint

$$E \leq \int_{\omega \in \Omega} p(\omega) y(\omega) d\omega.$$

Therefore, in general, demand for each variety ω is given by

$$y(\omega) = A p(\omega)^{-\frac{1}{1-\alpha}} I(\omega)^{\frac{1}{1-\alpha}}, \tag{2}$$

where

$$A = \frac{E}{\int_{\omega \in \Omega} p(\omega)^{-\frac{\alpha}{1-\alpha}} I(\omega)^{\frac{1}{1-\alpha}} d\omega}.$$

From equation (2) we can see how demand responds to a boycott. In this case the indicator variable takes the value of zero and there is no demand for the product.[14] The value of the preference shifter depends on the choice of the firm and nature (in the baseline model) or the activity of an advocacy NGO (in the microfounded model). This stylized assumption allows us to generate the risk of losing final revenues as a consequence of unethical production in a simple way that preserves tractability of the

in quality rather than productivity.

[14] Technically, this modeling is a variation of the standard approach in the literature on quality and international trade with CES preferences, where firms can invest into quality represented by a (usually continuous) variable which takes the place of our indicator variable. See e.g. Hummels and Klenow (2005) and Hallak (2006) for early contributions.

model.[15]

2.1.2 Production of the Final Good and the Intermediate Input

The final good is produced by the headquarter located in the Global North using an intermediate good provided by the supplier located in the Global South. The headquarter can costlessly transform one unit of an intermediate good into final output:

$$y(\omega) = x(\omega). \tag{3}$$

The quantity $y(\omega)$ produced of the final good is therefore simply given by the quantity $x(\omega)$ of the intermediate good the headquarter has at its disposal. The intermediate good is in turn produced by a supplier combining a headquarter service and a manufacturing input according to the following production function:

$$x(\omega) = \left(\frac{h(\omega)}{\beta}\right)^{\beta} \left(\frac{m(\omega)}{1-\beta}\right)^{1-\beta} \tag{4}$$

where $\beta \in (0, 1)$ is the headquarter intensity of production. The headquarter service $h(\omega)$ is provided to the supplier by the headquarter which then combines it with the manufacturing input $m(\omega)$ to produce $x(\omega)$ units of the intermediate good. The intermediate good produced is entirely relationship-specific. Neither can the supplier sell $x(\omega)$ to any third party nor can the headquarter produce any of the final output without the intermediate good that is in the possession of the supplier.[16] We stress that the manufacturing input m stands for a bundle of factors of production used by the supplier. Among these are labor and physical capital, as well as human capital and materials. In addition, and crucially for our model, the supplier also incurs other expenditures, such as provisions for workplace safety and the cost of compliance with local environmental regulation in the process of providing the input m.

2.1.3 Unethical Production and Consumer Boycotts

The central innovation in this paper is that the supplier does not only choose the investment necessary to produce the manufacturing input, but can also choose between a high- and a low-cost technology. The low-cost technology produces a (higher) negative externality on a third party. We can think of such externalities as taking the form of exploitation of workers with forced overtime, low work safety standards or child labor as well as pollution of the environment, e.g. by dumping dangerous chemicals in rivers,

[15] There are interesting microeconomic and behavioral issues related to this, in particular the question to which extent consumers form expectations about (un)ethical production and adjust consumption accordingly (as in Krautheim and Verdier, 2016) and to which extent the preference for ethical consumption is endogenous to NGO activity: Nyborg (2011) shows that consumers can be willing to pay to *not* receive information in order to avoid a moral obligation to contribute voluntarily. We do not seek to contribute to answering these questions and simply model consumer preferences such that we obtain the main features relevant for our analysis maintaining tractability.

[16] A setting where the headquarter provides an input to the supplier who can combine it with its own input and then take the produced output 'hostage' may seem a little artificial. The same may be the case for a final good technology that costlessly transforms the intermediate into the final product. We chose this modeling approach in order to reproduce the original Antràs (2003) setting. Later contributions consider settings where headquarter and supplier each produce an intermediate. Both intermediates are then combined to produce the final output. This setup appears more natural but delivers the same results and mechanisms as the original setting. To make the comparison to Antràs (2003) as clear as possible, we stick to the original modeling. Moreover, it is of interest to note that recent work on factoryless manufacturing points at examples like Apple or Dyson, who do not own any manufacturing establishments at all (Bernard and Fort, 2015).

emitting substantial quantities of carbon dioxide or harvesting old growth rainforests. Consumers consider an unnecessarily high (but cost-saving) level of this externality as unethical. We define the marginal cost of the supplier's high-cost, ethical technology as c_m^e and the low-cost, unethical technology by $c_m^u = \mu c_m^e$, with $\mu \in (0, 1)$.

We do not take any *normative* stand on what an ethical or unethical technology is. This includes for example the debate on the desirability of a ban of child labor. We simply start our modeling from the observation that consumer boycotts are triggered by the perception of (some) consumers that firms act in an unethical way. Clearly, what is considered 'unethical' may depend on the historical context, income, culture, salience of specific issues in the public debate as well as alternative technologies.

As consumers cannot infer from the final product whether the unethical technology was used in production, unethical firms can potentially prevent consumers from learning about the type of the firm. While we assume that the technology used cannot be directly observed by consumers, some firm choices are observable, potentially leading consumers to believe that the firm is of the unethical type. In the baseline version of the model we introduce a simple link between observable choices (investments, quantities and prices) of the firm and the probability of facing a boycott: an unethical firm setting observables at values that are optimal for an ethical firm has a chance to pass as an ethical firm and faces a boycott with probability $1 - \gamma < 1$. Any firm deviating from the investments, quantities or prices of ethical firms faces a boycott with probability one.[17] In fact, we only need to impose this for investments, as conditional on identical investments, the same quantities and prices maximize profits of both firm types.[18] This implies that an unethical firm faces discontinuous demand being positive in expectation if and only if it chooses investments like an ethical firm. This leads unethical firms to mimic ethical firms and in equilibrium the levels of investment, quantities and prices do not reveal the type of the firm.

Using this reduced-form approach in the baseline model allows us to focus on the analysis of the international organization of production with unethical technologies, to derive our main results on the integration and technology decision, their interaction as well as empirical implications. However, the reduced-form approach leaves some questions open. What is the mechanism/the agent triggering a consumer boycott? Why is it triggered by a deviation from ethical firm choices? Should ethical firms adjust their investments in order to signal their type? To address these questions, we provide a microfoundation in Section 3, where an NGO observes firm choices (organization of production, investments, quantities, prices) and can determine the optimal choices of an ethical firm. When the NGO observes a firm that acts inconsistently with the use of the ethical technology it starts an investigation. If it finds the firm to be of the unethical type it initiates a boycott. We show that all the results of the baseline model remain qualitatively unchanged when the model is fully microfounded. As all the additional assumptions introduced in the microfoundation serve the sole purpose of microfounding the link between mimicking and boycotts, but do not add major insights to our research questions, we keep them separated from the

[17] It is merely for tractability that we consider a setting where *any* deviation of the ethical investment triggers a consumer boycott. Even if this assumption were to be relaxed, the magnification of the outsourcing incentive through unethical production presented below should remain active as long as the supplier needs to stay below the level of investment it would optimally choose in the absence of the threat of a boycott.

[18] This is because after investments are made, the 'optimization' of a firm with respect to quantities and prices is equivalent to a situation in which all costs are sunk, marginal costs are zero, and the maximum output is fixed and identical for both firm types as investments are the same. Therefore, when both firms have set the same investment levels and there is positive demand for the unethical firm, both firms will set the same quantities and market clearing prices.

baseline model.

2.1.4 Hold-up Problem and the Organization of the Firm

We consider an environment with incomplete contracts. Neither can contracts be written contingent on choices the parties make, nor on outcomes like revenue. The only contractible items are the lump-sum transfer from the supplier to the headquarter (discussed in detail below) and the organization of production. This means that investment quantities are not contractible, but also that our new feature, the technology choice of the supplier, cannot be contracted upon.[19]

As contracts are incomplete, neither the investments nor the split of the revenues can be fixed ex-ante. The relationship-specificity of investments then implies that after investments are sunk and the intermediate input is produced, the two parties face a hold-up problem. Both parties need the partner to generate (full) revenue and therefore engage in a bargaining process over the split of final revenues. Following the literature, we model this ex-post bargaining as generalized Nash bargaining with the headquarter getting a fraction of the final revenues. This fraction is endogenous and depends on the residual rights of control, which are in turn affected by the organization of production chosen by the headquarter.

Before investments take place, the headquarter can choose between integrating the supplier into the firm or leaving it as an independent party. We index the mode of organization by $k \in \{O, V\}$, where O stands for international outsourcing and V for vertical integration. The key difference between the two is that outsourcing leaves the supplier with the residual rights of control over the produced intermediate. In this case the outside options of both parties are zero if bargaining fails: the headquarter has no input to produce the final product and the supplier cannot transform the intermediate into the final product.[20] Integration in turn shifts the residual rights of control to the headquarter allowing it to recover a fraction $\delta \in (0,1)$ of the intermediates from the supplier if bargaining fails. The outside option of the headquarter under integration is therefore better than under outsourcing, implying that the bargaining results in a larger share of revenues going to the headquarter, i.e. $\phi_V > \phi_O$, where, as in Antràs (2003), $\phi_V = \phi_O + \delta^{\alpha}(1 - \phi_O)$.

We assume $\phi_k > \frac{1}{2}$. Antràs (2003) shows that this assumption is sufficient to ensure that the headquarter optimally produces the headquarter service by itself and hands it over to the supplier for production of intermediate $x(\omega)$ while the supplier produces the manufacturing input.[21]

[19] Alternatively, one could assume that integration allows the headquarter to *impose* the technology on the supplier. This would, however, mix property rights theory (for production) and the transaction cost approach à la Grossman and Helpman (2002) or Carluccio and Bas (2015) (for technology). In the latter, all contractual incompleteness is resolved by integration. It appears hard to justify the assumption that under integration the headquarter *can* impose the type of technology but *cannot* impose the level of investment. More interesting might be the analysis of a setting that fully embraces the logic of the transaction cost approach where integration allows the headquarter to impose both the investment and the technology. We leave this alternative model for future research and focus in this paper on the predominant paradigm in the literature: the property-rights theory of the firm.

[20] The assumption that the supplier cannot get anything out of its residual rights of control can easily be relaxed e.g. by allowing the supplier to sell the intermediate good at a discounted rate on a secondary market. This does not affect the qualitative results (see e.g. Antràs and Yeaple, 2014).

[21] This assumption implies that we are considering a two-sided hold-up problem, where both parties have sunk an investment in their specific factor. This assumption is therefore key to establish the qualitative equivalence to setups briefly outlined in Footnote 16 where the respective inputs are only combined after bargaining was successful.

2.1.5 Match Creation and Transfer Payment

We have now described the situation after a headquarter has been matched to a supplier. Ex-ante, the headquarter faces a large number of perfectly competitive suppliers available for a match. Once a match is formed, their relationship is transformed into one of bilateral monopoly (Williamson, 1985) in that investments are relationship-specific and have no outside value. Due to incomplete contracts, the production process involving bargaining over the revenues will leave the supplier with positive profits. The large number of potential suppliers compete for this profitable opportunity by offering a transfer payment to the headquarter in return for forming the match with them. Perfect competition among suppliers implies that the headquarter can set a payment that extracts the full expected surplus from the supplier. Besides the organization of production, the transfer payment is the only variable the headquarter and supplier can contract on. Both are fixed in the moment the match is formed.

2.1.6 Time Line

Figure 1 gives an overview of the sequence of events. In t_0, the headquarter chooses the organizational form and the lump-sum transfer. In $t_1(a)$, the supplier chooses between ethical and unethical production. Both parties make their physical investments non-cooperatively in $t_1(b)$. The headquarter hands the headquarter service to the supplier, who in turn produces intermediate inputs in t_2 by combining the headquarter service with its own manufacturing input. In t_3, nature determines whether an unethical firm will be boycotted by consumers. Period t_4 features the ex-post bargaining over the division of the surplus. In t_5, if the parties have agreed on a division, intermediates are converted to final output, sold and revenues distributed to headquarter and supplier if the firm is not boycotted. In case of a boycott, demand is zero and no final goods are produced and sold.

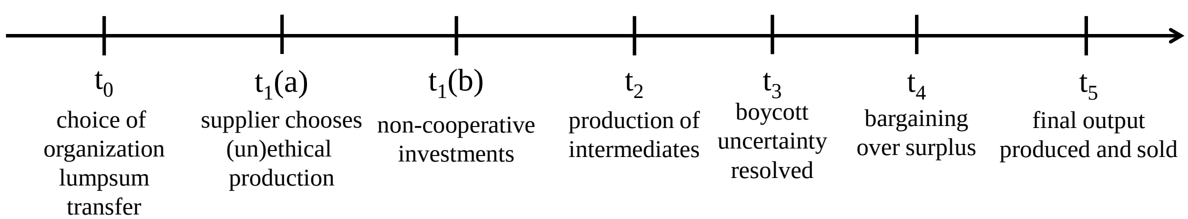

Figure 1: Timing of events.

2.2 Equilibrium Firm Choices

We solve the model by backward induction.

2.2.1 t_5: Revenues of Ethical and Unethical Firms

We denote revenue from selling variety ω as $R(\omega)_k^l$, where $k \in \{V, O\}$ indicates vertical integration and outsourcing and $l \in \{e, u\}$ indicates ethical and unethical production. An ethical firm always faces full demand as it is never targeted by a consumer boycott. Its revenues are given by $R(\omega)_k^e = p(\omega)_k^e\ y(\omega)_k^e$. $h(\omega)_k^e$ and $m(\omega)_k^e$ represent the investment quantities chosen by headquarter and supplier in the case of ethical production. Given that the quantity $x(\omega)$ of the intermediate good produced by the supplier is determined by investments and given that the headquarter costlessly transforms $x(\omega)$ into $y(\omega)$, total revenues of an ethical firm can be expressed as

$$R(\omega)_k^e = A^{1-\alpha}\left[\left(\frac{h(\omega)_k^e}{\beta}\right)^{\beta}\left(\frac{m(\omega)_k^e}{1-\beta}\right)^{1-\beta}\right]^{\alpha}. \tag{5}$$

An unethical firm only faces positive demand in expectation if $h(\omega)_k^u = h(\omega)_k^e$ and $m(\omega)_k^u = m(\omega)_k^e$, its revenues under mimicking and if it does not face an exogenous boycott in t_3 are also given by the above expression.

2.2.2 t_4: Bargaining

Headquarter and supplier bargain over the distribution of revenue. The bargaining power - and therefore also the share of revenue - of the headquarter is assumed to be $\phi_O > \frac{1}{2}$ under outsourcing. This reflects the fact that in the arm's length relationship, both parties have an outside option of zero and the payoff allocation is determined only by the exogenous assumptions about the distribution of the gains from trade. In the case of integration, the outside option of the supplier remains at zero because of the relationship-specificity of the produced intermediates. The headquarter, however, has allocated the residual rights of control to itself. It is able to continue producing $\delta y(\omega)$ in case bargaining breaks down. Using equations (3), (4), and (5) this translates into sales of $\delta^{\alpha} R(\omega)_k^l$. The gains from trade are thus reduced to $(1-\delta)^{\alpha}\, R(\omega)_k^l$. With integration, the headquarter receives its larger outside option plus its exogenous share from the gains from trade, which is $\phi_V R(\omega)_k^l$, with ϕ_V as defined in Section 2.1.4.

2.2.3 t_3 and t_2: Consumer Boycotts and Production of Intermediates

In period t_3 nature decides whether an unethical firm faces a boycott. We assume that ethical firms never face a boycott, firms that are openly unethical always face a boycott and firms that mimic ethical firms in terms of prices, output, and investment face a boycott with a probability $1-\gamma$. Before the boycott uncertainty is resolved, a mimicking unethical firm therefore has an expected revenue of

$$E[R(\omega)_k^u] = \gamma R(\omega)_k^e. \tag{6}$$

In period t_2, the supplier uses the invested quantities to produce intermediate output $x(\omega)$. As outlined above, provided it mimicked in terms of investments in t_1, there is no reason for an unethical firm to deviate from the optimal quantity of an ethical firm, which is production according to equation (4).

2.2.4 $t_1(b)$: Investments

Two types of decisions are taken sequentially in period t_1. In period $t_1(a)$ the supplier chooses to implement the ethical or unethical technology. In period $t_1(b)$ supplier and headquarter take their investment decisions simultaneously. We first consider the investment choices conditional on the ethical or unethical technology being implemented.

Ethical Investments: When the supplier implements the ethical technology, the setting is isomorphic to Antràs (2003). The two parties simultaneously and non-cooperatively set investments to maximize their respective shares of final revenue. They take into account incomplete contracts and the ensuing ex-post bargaining. The headquarter maximizes

$$\max_{h(\omega)_k^e} \phi_k R(\omega)_k^e - c_h h(\omega)_k^e, \tag{7}$$

whereas the supplier solves

$$\max_{m(\omega)_k^e} (1-\phi_k)\, R(\omega)_k^e - c_m^e m(\omega)_k^e. \tag{8}$$

Notice the superscript in the marginal cost of the supplier. With ethical production, the supplier rewards its factor of production at the ethical rate c_m^e.

The first order conditions deliver the best response functions that give optimal investment of each party for any positive level of investment of the other party:

$$h(\omega)_k^e = \beta \left(\frac{\phi_k \alpha}{c_h}\right)^{\frac{1}{1-\beta\alpha}} A^{\frac{1-\alpha}{1-\alpha\beta}} \left(\frac{m(\omega)_k^e}{1-\beta}\right)^{\frac{(1-\beta)\alpha}{1-\beta\alpha}}$$

$$m(\omega)_k^e = (1-\beta)\left(\frac{(1-\phi_k)\,\alpha}{c_m^e}\right)^{\frac{1}{1-(1-\beta)\alpha}} A^{\frac{1-\alpha}{1-(1-\beta)\alpha}} \left(\frac{h(\omega)_k^e}{\beta}\right)^{\frac{\beta\alpha}{1-(1-\beta)\alpha}}.$$

Curve S_V in the left panel of Figure 2 depicts the supplier's best response function, H_V the headquarter's best response function under vertical integration, S_O and H_O do the same for outsourcing. S^* and H^* show the best responses in the first best case, which is unattainable because of incomplete contracts.

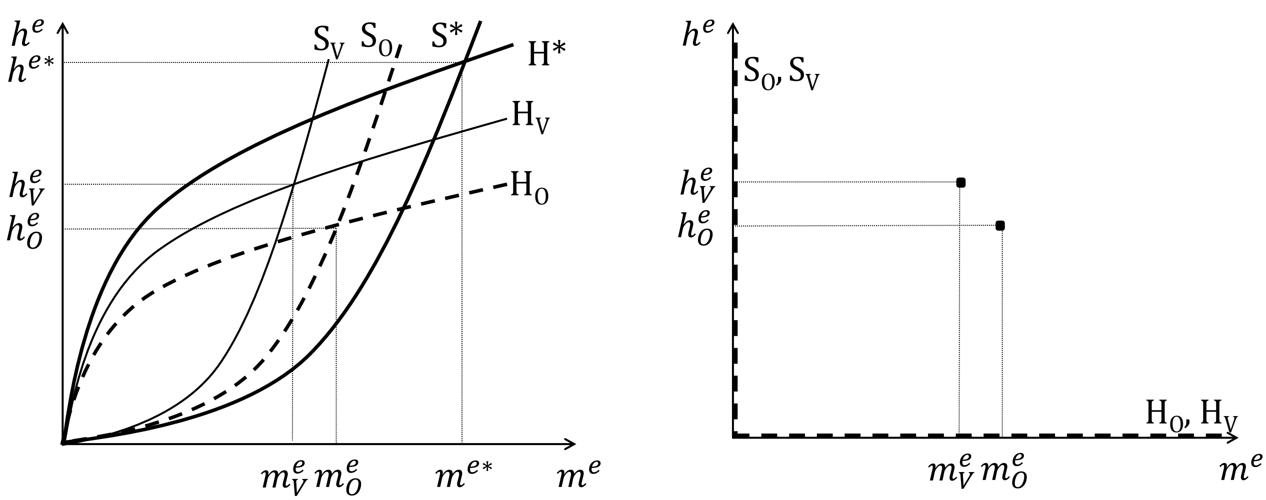

Figure 2: Best response functions under ethical (left) and unethical production (right).

Like in Antràs (2003), the equilibrium of the investment game is at the intersection of the best response functions. The standard argument of Pareto-dominance rules out the other Nash equilibrium at zero-zero.

Equilibrium investments are therefore given by

$$h(\omega)_k^e = \beta A \alpha^{\frac{1}{1-\alpha}} \frac{\phi_k}{c_h} \left[\left(\frac{c_h}{\phi_k} \right)^{\beta} \left(\frac{c_m^e}{1-\phi_k} \right)^{1-\beta} \right]^{\frac{-\alpha}{1-\alpha}} \tag{9}$$

$$m(\omega)_k^e = (1-\beta) A \alpha^{\frac{1}{1-\alpha}} \frac{1-\phi_k}{c_m^e} \left[\left(\frac{c_h}{\phi_k} \right)^{\beta} \left(\frac{c_m^e}{1-\phi_k} \right)^{1-\beta} \right]^{\frac{-\alpha}{1-\alpha}}. \tag{10}$$

We refer to these investments as the *baseline ethical investment profile* $i(\omega)_k^* = \{h(\omega)_k^e, m(\omega)_k^e\}$. Plugging (9) and (10) into revenue from (5) gives equilibrium revenue generated by an ethical firm as

$$R(\omega)_k^e = A \alpha^{\frac{\alpha}{1-\alpha}} \left[\left(\frac{c_h}{\phi_k} \right)^{\beta} \left(\frac{c_m^e}{1-\phi_k} \right)^{1-\beta} \right]^{\frac{-\alpha}{1-\alpha}}. \tag{11}$$

Unethical Investments: We now turn to the non-cooperative investment game when the supplier has chosen the unethical technology. Demand is still given by equation (2), but the difference is that the indicator variable $I(\omega)$ may also take the value of zero. This is the case when the unethical firm does not mimic or if it faces an exogenous boycott in t_3. Mimicking involves setting the same price as the ethical firm. Therefore, the demand function becomes degenerate. When the unethical firm sets the ethical price, $I(\omega) = 1$ and it gets full demand. As soon as it deviates from it, we have $I(\omega) = 0$ and therefore zero demand.

An ethical firm faces a continuous demand function, leading to the continuous best response functions derived above. Consider the case that an unethical supplier would prefer mimicking over zero production. This is the only relevant case, as otherwise no supplier would choose unethical production in the first place. In this case the best response functions for the unethical firm are symmetric for the headquarter and the supplier and are given by

$$h(\omega)_k^u = \begin{cases} h(\omega)_k^e & \text{if } m(\omega)_k^u = m(\omega)_k^e \\ \textit{undetermined} & \text{if } m(\omega)_k^u = 0 \\ 0 & \textit{otherwise} \end{cases}$$

$$m(\omega)_k^u = \begin{cases} m(\omega)_k^e & \text{if } h(\omega)_k^u = h(\omega)_k^e \\ \textit{undetermined} & \text{if } h(\omega)_k^u = 0 \\ 0 & \textit{otherwise.} \end{cases}$$

The best response functions are illustrated in Figure 2. Different to the ethical case, they take a value of zero for *any* investment of the other party deviating from the baseline ethical investment (indicated by the bold dashed lines). The only point with positive investments of both parties is when they both set the baseline ethical investment.

While the best response functions are fundamentally different from the ones for the ethical firm, they share the Nash equilibria at zero-zero and the baseline ethical investments. In fact, they lead to the same equilibrium of the investment game. To see this, note that no party would ever find it optimal to choose an investment that is not on its best response function, as it would be strictly dominated by playing the best response. This implies that only two investments can occur for each party: zero or the baseline ethical investment. As in the case with ethical production we invoke the Pareto-dominance criterion so that the equilibrium with positive investment is the one that is played.[22]

2.2.5 $t_1(a)$ (Un)ethical technology choice

We have seen how the non-cooperative investment decisions are taken for ethical and unethical firms in period $t_1(b)$. Based on this, we can now turn to period $t_1(a)$ analyzing the supplier's choice between the two technologies. In taking the technology decision, the supplier faces a trade-off between the cost savings implied by unethical production and the risk of losing its share of total revenues due to a consumer boycott.

First consider the determinants of the expected revenues of the supplier. A (mimicking) unethical firm still faces a boycott with probability $1-\gamma$ so that expected revenues are given by $E[R(\omega)_k^u] = \gamma R(\omega)_k^e$. With a fraction $1-\phi_k$ going to the supplier and given the equilibrium $R(\omega)_k^e$ in equation (11), expected revenues of an unethical supplier are given by

$$(1-\phi_k)E\left[R(\omega)_k^u\right] = \gamma(1-\phi_k)A\alpha^{\frac{\alpha}{1-\alpha}}\left[\left(\frac{c_h}{\phi_k}\right)^{\beta}\left(\frac{c_m^e}{1-\phi_k}\right)^{1-\beta}\right]^{-\frac{\alpha}{1-\alpha}}. \tag{12}$$

The expected difference between ethical and unethical revenues of the supplier is

$$E[\Delta R_S] = (1-\phi_k)\left(R(\omega)_k^e - E\left[R(\omega)_k^u\right]\right).$$

This difference is always positive and reflects the fact that ethical firms have higher revenues in expectation, as they always face full demand. We refer to this difference as the *ethical revenue premium*. The supplier trades off this ethical revenue premium against the cost savings of unethical production. The unit cost savings are determined by the scaling factor $\mu = \frac{c_m^u}{c_m^e}$ where $1-\mu \in (0,1)$ can be interpreted as the unit cost savings of unethical production which we refer to as the *unethical cost advantage*. Total cost savings of unethical production are given by $\Delta C = (c_m^e - c_m^u)\, m(\omega)_k^e$. With $m(\omega)_k^e$ given by equation (10).

In stage t_1, the organizational decision as well as the lump-sum transfer are fixed, as they are set in t_0. The supplier therefore takes the decision on unethical production by trading off $E[\Delta R_S]$ against ΔC. This decision can be described by a cutoff headquarter intensity β_S above which the supplier chooses the ethical technology and below which it produces unethically.

Proposition 1 *The headquarter intensity of a sector influences the technology choice of the supplier.*

[22] An alternative way to rationalize the equilibrium with positive investments would be to assume that investments become relationship-specific if and only if both sides make a positive investment. So as soon as both sides make a positive investment, all the properties of the baseline model apply. But in the case in which one party makes zero investment, the input remains 'pure' and can be resold on the factor market at zero cost. Intuitively, this technology works like mixing red and white liquid paint. Two parties non-cooperatively decide the quantity of their type of paint they put into the same bucket. Once mixed, both inputs cannot be recovered. But in the special case where zero of the red paint is added, the white paint is not contaminated (not match-specific) and can be resold on the factor market for white paint (and vice versa).

Specifically, the supplier chooses unethical production when the headquarter intensity β is lower than

$$\beta_S = 1 - \frac{1-\gamma}{\alpha\,(1-\mu)}. \tag{13}$$

The cutoff β_S (i) increases in the unethical cost advantage, $\frac{\partial \beta_S}{\partial (1-\mu)} > 0$; (ii) decreases in the probability of a boycott, $\frac{\partial \beta_S}{\partial (1-\gamma)} < 0$; (iii) and decreases in the mark-up, $\frac{\partial \beta_S}{\partial (1/\alpha)} < 0$.
Proof: *See the Appendix.*

Proposition 1 implies a direct link between headquarter intensity and (un)ethical production. Firms in sectors with a high supplier (low headquarter) intensity tend to implement the unethical technology, while ethical production is more likely in headquarter-intensive sectors. The choice between ethical and unethical production is driven by the trade-off between the supplier's total cost savings of unethical production and the supplier's expected loss of final revenue through a potential boycott. First, note that a high supplier intensity (low β) scales up the supplier's investment and therefore the potential cost savings from unethical production. Therefore, the unethical technology tends to be implemented in the supplier intensive sectors. This split of sectors into ethical and unethical ones is also affected by the other variables in equation (13).

At given investments, a stronger unethical cost advantage (the unit cost savings of unethical production $1-\mu$) scales up total cost savings and makes unethical production attractive also for suppliers with lower levels of investments (i.e. in more headquarter intensive industries). The probability of facing a boycott $1-\gamma$ reduces expected revenues of the supplier under unethical production and therefore makes it less attractive. The mark-up a firm charges over its marginal cost is given by $1/\alpha$. A higher mark-up (strong market power) implies higher profits *per unit*. As (the supplier's share of) these profits are traded off against *per unit* cost savings, it unambiguously discourages unethical production. This implies that firms in sectors with high market power and large profit margins, which have more to lose from a boycott, tend to use ethical technologies.

Corollary 1 *The supplier's choice between the ethical and unethical technology is independent of the bargaining power and is therefore not affected by the organization of production (outsourcing vs. integration).*
Proof: *Simply note that the organization only affects the bargaining power of the headquarter and the supplier. It follows from equation* (13) *that the choice between ethical and unethical technology is independent of the bargaining power and does therefore not depend on the organization of production.*

The fact that the bargaining power and therefore the organization of production does not affect the choice between ethical and unethical production has an important implication in our model. We have seen in Section 2.1.4 that, by choosing between integration and outsourcing, the headquarter can affect the bargaining power and thereby the investments of the two parties. The organization of production therefore provides an instrument for the headquarter to affect the non-contractible investment choice of the supplier. Corollary 1 implies, however, that this is no instrument the headquarter can use to influence the technology choice of the supplier: the decision for or against unethical production is independent of the bargaining power and is therefore also independent of the organization of production.

The reason for this is that it affects the technology decision through two opposing effects offsetting each other. On the one hand, a stronger bargaining power increases the share of total revenue going to the supplier. This increases the losses in case of a boycott and incentivizes ethical production. On the other hand, by increasing the share of total revenues, the higher bargaining power also increases the optimal investment level. This scales up the cost savings of unethical production. The derivation of equation (13) in the Appendix shows that the two effects exactly offset each other.

2.2.6 t_0: Optimal Organizational Structure and Transfer Payment

Transfer Payment Taking into account incomplete contracts, the investments in the manufacturing input and the equilibrium outcome of the ex-post bargaining a supplier in a sector in which $\beta > \beta_S$ knows its private profits are going to amount to

$$\pi^e_{k,S} = (1 - \phi_k)\, R(\omega)^e_k - c^e_m m(\omega)^e_k \tag{14}$$

if it enters the match with the headquarter which has chosen organizational form $k \in \{O, V\}$. In the other case, in which a supplier knows it will choose unethical production and mimicking because $\beta < \beta_S$, it expects to earn

$$E\left[\pi^u_{k,S}\right] = \gamma\,(1 - \phi_k)\, R(\omega)^e_k - c^u_m m(\omega)^e_k \tag{15}$$

in case of a successful match. Because the headquarter faces a large number of potential suppliers competing perfectly for the opportunity to produce the final good with it, these private profits represent the maximum amount a supplier is willing to pay for this opportunity. The headquarter knows its own β and has decided the optimal organizational form $k \in \{O, V\}$. Given this decision and anticipating the technology choice of the supplier in t_1 the headquarter extracts

$$T_k = \begin{cases} \pi^e_{k,S} & \text{if } \beta > \beta_S \\ E\left[\pi^u_{k,S}\right] & \text{if } \beta < \beta_S. \end{cases} \tag{16}$$

Organizational choice At the same time, the headquarter chooses between integration and outsourcing maximizing the total surplus of the match. Both decisions depend on the supplier's anticipated technology choice in stage t_1.

As the supplier's choice of technology does not depend on the bargaining power ϕ_k, the headquarter observes the headquarter intensity of its sector and perfectly foresees the technology choice of the supplier. Therefore, in the case of $\beta > \beta_S$, the headquarter anticipates ethical production by the supplier. In this case the total surplus of the match is given by the sum of the two parties' private profits

$$\Pi^e_k = R(\omega)^e_k - c^e_m m(\omega)^e_k - c_h h(\omega)^e_k. \tag{17}$$

If $\beta < \beta_S$, the headquarter knows the supplier will choose the unethical technology and mimic an ethical firm in investments, quantities and prices. The total surplus of the match is then subject to the uncertainty generated by the threat of a consumer boycott and is given by

$$E\left[\Pi^u_k\right] = \gamma R(\omega)^e_k - c^u_m m(\omega)^e_k - c_h h(\omega)^e_k. \tag{18}$$

In deciding the organizational form of the firm the headquarter compares the overall value of the relationship under outsourcing to the overall value under integration taking the technology choice of the supplier as given. Given ethical production by the supplier, the ratio of total profits under integration and total profits under outsourcing is given by

$$\Theta^{e}(\beta)=\left[\left(\frac{\phi_V}{\phi_O}\right)^{\beta}\left(\frac{1-\phi_V}{1-\phi_O}\right)^{1-\beta}\right]^{\frac{\alpha}{1-\alpha}}\frac{1-\alpha\left(1-\beta\right)+\phi_V\alpha\left[1-2\beta\right]}{1-\alpha\left(1-\beta\right)+\phi_O\alpha\left[1-2\beta\right]}.$$

The cutoff headquarter intensity above which the headquarter offers to the supplier a contract stipulating integration of the supplier and the transfer payment T_V given that it produces ethically ($\beta > \beta_S$) is implicitly defined by

$$\Theta^{e}(\beta_e)=1. \tag{19}$$

Given unethical production by the supplier, the ratio of total expected profits is given by

$$\Theta^{u}(\beta)=\left[\left(\frac{\phi_V}{\phi_O}\right)^{\beta}\left(\frac{1-\phi_V}{1-\phi_O}\right)^{1-\beta}\right]^{\frac{\alpha}{1-\alpha}}\frac{\gamma-\alpha\left(1-\beta\right)\mu+\phi_V\alpha\left[\mu-\beta\left(1+\mu\right)\right]}{\gamma-\alpha\left(1-\beta\right)\mu+\phi_O\alpha\left[\mu-\beta\left(1+\mu\right)\right]}.$$

The cutoff headquarter intensity β^u above which the headquarter offers to the supplier a contract stipulating integration of the supplier and the transfer payment T_V given that it produces unethically ($\beta < \beta_S$) is implicitly defined by

$$\Theta^{u}(\beta_u)=1. \tag{20}$$

The expression differs from $\Theta^e(\beta_e)$ in two respects. Because of unethical production there is now a threat of a boycott and second, the unethical cost advantage is exploited by the supplier. We summarize our result in the following subsection.

2.3 (Un)ethical Production, Factor Intensity and Ownership Structure

We can now combine the above insights on the implementation of the (un)ethical technology and the organizational choices of the firm conditional on technology to analyze the equilibrium of the model. Most notably, we are interested in the question of how the technology choice of the supplier interacts with the integration decision of the headquarter.

2.3.1 The Unethical Outsourcing Incentive

Based on equations (19) and (20), we can state the following proposition:

Proposition 2 *There exists a unique β_e below which the headquarter chooses outsourcing irrespective of the technology choice of the supplier. Integration is always chosen for headquarter intensities above β_u and it always holds that $\beta_e < \beta_u$. A sufficient condition for a unique interior solution $\beta_u \in (\beta_e, 1)$ to exist is given by $\gamma > \frac{4\phi_V}{3+\phi_V}$. For any $\beta \in (\beta_e, \beta_u)$ the headquarter chooses integration if and only if the supplier produces ethically and chooses outsourcing if and only if unethical production is anticipated.*
Proof: *See the Appendix.*

The parameter condition $\gamma > \frac{4\phi_V}{3+\phi_V}$ is sufficient to ensure that $\beta_u < 1$ implying that both outsourcing and integration are chosen for some levels of headquarter intensity. Since we are interested in the interaction of unethical production with the organization of production, we focus on the cases in which both types of organizational form can emerge. However, $\beta_e < \beta_u$ regardless of whether the above condition holds.

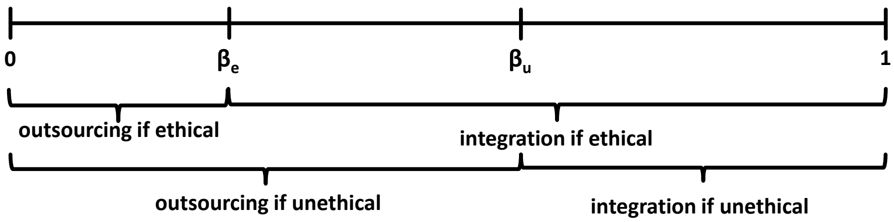

Figure 3: Unethical production and the two integration cutoffs.

Figure 3 highlights the pattern described in Proposition 2. The axis shows the range of admissible headquarter intensities implying high supplier intensity on the left and high headquarter intensity on the right. The cutoff β_e is identical to the cutoff in Antràs (2003). It reflects the fact that the headquarter faces two underinvestment problems in period t_1 (the headquarter's and the supplier's). The organization of production is an instrument to alleviate the underinvestment of either the headquarter (through integration) or the supplier (through outsourcing). The mechanism is that integration and outsourcing imply different residual rights of control for the headquarter and the supplier. This changes the bargaining power and thereby the share of total revenue each party obtains. As a larger share of revenue increases the optimal investment, integration alleviates the headquarter's underinvestment while outsourcing alleviates the supplier's underinvestment. We refer to this pattern as the *Antràs implication*.

When the supplier chooses unethical production, the attractiveness of outsourcing increases above and beyond the Antràs implication: unethical production reduces the unit costs of the manufacturing input so that the difference between the actual and the optimal investment increases. This aggravates the underinvestment problem of the supplier compared to the case of ethical production with the same headquarter intensity. The headquarter responds to this by expanding the use of the now cheaper manufacturing input as much as possible. It can achieve this by shifting the residual rights of control to the supplier through outsourcing to incentivize a larger ex-ante investment. We call this the *unethical outsourcing incentive*. It is captured by the cutoff β_u. The fact that $\beta_e < \beta_u$ shows that outsourcing is chosen by the headquarter for a larger range of headquarter intensities if the supplier produces unethically. In particular, the *unethical outsourcing incentive* distorts the *Antràs implication* towards outsourcing so that the headquarter chooses outsourcing solely *because of* unethical production for $\beta \in (\beta_e, \beta_u)$. This implies that the supplier's technology choice can affect the organizational choice of the headquarter. Specifically, the headquarter tends to keep unethical suppliers at arm's length.

2.3.2 Ethical Integration and Unethical Outsourcing?

The equilibrium pattern of (un)ethical production and the organization of production depends on how the cutoffs β_S, β_e and β_u relate to one another. The following proposition summarizes the relevant cases to be distinguished.

Proposition 3 *There exist three possible equilibria of the model characterized by $\beta_e < \beta_S < \beta_u$ (Case 1); $\beta_e < \beta_u < \beta_S$ (Case 2) and $\beta_S < \beta_e < \beta_u$ (Case 3). Unethical outsourcing and ethical integration are*

equilibrium outcomes in all three cases. Unethical integration and ethical outsourcing can also occur in equilibrium in Cases 2 and 3, respectively.
Proof: *See the Appendix.*

Proposition 3 implies that unethical production and outsourcing are associated in our model as are ethical production and integration. The reason is that the cost savings of unethical production are scaled by the size of the supplier's investment, which is larger in sectors with high supplier intensity (lower headquarter intensity) of production. At the same time, the Antràs mechanism implies that sectors with a high supplier intensity optimally shift bargaining power to the supplier through outsourcing to mitigate the underinvestment problem where it is most severe. Taken together, sectors with high supplier intensities tend to implement outsourcing and unethical production, while sectors with a high headquarter intensity tend to feature ethical production and integration. This finding is consistent with the claim by Mosley (2011) that outsourcing is connected with a lowering of labor standards, while integrated multinationals tend to raise them. Our model is rich enough, however, to also feature ethical outsourcing and unethical integration as equilibrium outcomes.

This is illustrated in Figure 4. In Case 1, β_S is in between β_e and β_u. In this case the cutoff splitting sectors into ethical and unethical ones also splits the sectors into integrating and outsourcing ones. Cases 2 and 3 illustrate what happens if the attractiveness of unethical production is very strong or very weak (e.g. because of the cost advantage of unethical production analyzed in detail below). In the former case (Case 2), unethical production is so attractive that the headquarter decides to integrate despite the use of the unethical technology by the supplier. In the latter case (Case 3), ethical outsourcing occurs for a range of headquarter intensities. This illustrates that there is no mechanical link between outsourcing and unethical production in our model. Whether unethical integration and ethical outsourcing occur depends on the strength of the Antràs mechanism, the attractiveness of unethical production as well as the importance of the unethical outsourcing incentive.

2.3.3 Incentives for (Un)ethical Production and the Organization of the Firm

We are ultimately interested in the question if and how the (un)ethical technology choice of the supplier interacts with the organization of production. To address this question, we define $\bar{\beta}$ as the headquarter intensity above which integration actually takes place. This cutoff is given by $\bar{\beta} = \beta_S$ in Case 1; $\bar{\beta} = \beta_u$ in Case 2; and $\bar{\beta} = \beta_e$ in Case 3. With $\beta_e < \beta_u$, we can write the integration cutoff as:

$$\bar{\beta} = \begin{cases} \min\{\beta_S; \beta_u\} & \text{if } \beta_S > \beta_e \\ \beta_e & \text{otherwise.} \end{cases} \tag{21}$$

Proposition 4 *The outsourcing cutoff is weakly increasing in the unethical cost advantage, i.e.* $\frac{\partial \bar{\beta}}{\partial(1-\mu)} \geq 0$.
Proof: *See the Appendix.*

We can see from Proposition 4 that the outsourcing cutoff is weakly increasing in the unethical cost advantage parameter $1 - \mu$. An increase in the unethical cost advantage increases both β_S and β_u and

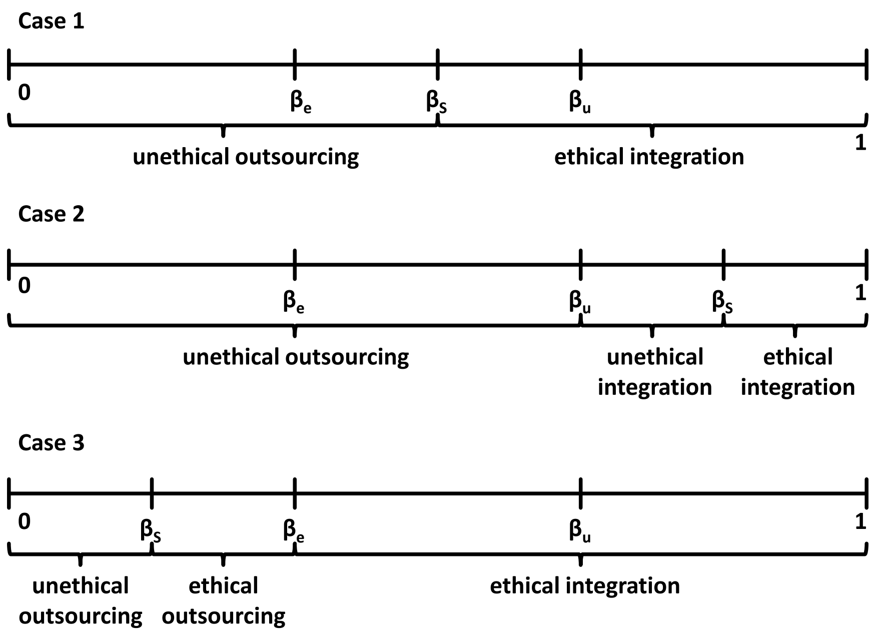

Figure 4: Interaction of unethical production and the outsourcing decision.

when unethical production surpasses a minimum level of attractiveness for the supplier ($\beta_e < \beta_S$), this unambiguously increases the integration cutoff $\bar{\beta}$. This implies that besides the variables that affect β_e that have already been accounted for in the literature, our model identifies the unethical cost advantage as a new parameter that affects the integration decision of the firm. We will exploit this implication in our empirical analysis in Section 4. In this context $1-\mu$ has an industry and a country dimension, consisting of an industry-specific (technology driven) cost savings potential and country-specific differences in regulation and enforcement capacity.

The intuition behind the above result is as follows. Case 3 represents the case where unethical production is very unattractive. In this case a marginal change in $1-\mu$ does not affect outsourcing. Consider the case where there is no unethical cost advantage at all ($1-\mu = 0$). In this case unethical production is never optimal for the supplier and outsourcing is determined by the Antràs mechanism only. When we increase $1-\mu$, the least headquarter-intensive industries start to use the unethical technology, but they are under the outsourcing regime anyway, so that the unethical outsourcing incentive does not alter the policy of the headquarter.

Once $1-\mu$ is large enough to have $\beta_e < \beta_S$ the picture changes. In this case the unethical outsourcing incentive makes firms opt for outsourcing that would otherwise choose integration. As both β_S and β_u increase in $1-\mu$, outsourcing increases in $1-\mu$ both in Case 1 and Case 2. The cutoff β_S represents the incentives for unethical production for the supplier while β_u reflects the optimal response to it by the headquarter. As a stronger cost advantage makes unethical production more attractive, β_S increases in $1-\mu$.

For β_u, note that when the headquarter anticipates unethical production, the damage is done (in expectation) on the demand side: a boycott occurs and reduces demand to zero with probability $1-\gamma$. As the headquarter can influence neither the technology decision nor the effect of a boycott, it takes these as given and has an incentive to maximize the benefits of unethical production by increasing the supplier's manufacturing investment through outsourcing. This is the *unethical outsourcing incentive* discussed above. A higher cost advantage of unethical production therefore increases the range of headquarter intensities for which outsourcing is chosen by the headquarter, i.e. β_u increases.

2.3.4 Headquarter's perspective on ethical production: aspirations and reality

Before we proceed to analyzing the microfoundation of the link between the boycott and prices, output and investments and before we present empirical test of Proposition 4 in the following sections, we now highlight an interesting tension that arises between the headquarter's aspirations and actions regarding (un)ethical production. Consider a headquarter that states that it would like to source its products ethically but then incentivizes its suppliers to expand unethical production. An external observer may interpret this as evidence of a dishonest attempt of greenwashing or - simply put - a lie by the firm. Our model, however, implies that this combination of actually *wishing* to source ethically but *expanding* unethical production can be an equilibrium outcome.

For this situation to occur two conditions have to be met. First, we need to be in a situation where the headquarter chooses outsourcing if and only if unethical production is anticipated, i.e. $\beta \in (\beta_e, \beta_u)$ (condition 1). We have seen in the discussion of Proposition 2 that in this range the only reason to opt for outsourcing rather than integration is to expand unethical production to fully benefit from the unethical cost advantage. Second, within this range there must be a non-empty set of headquarter intensities for which the headquarter would impose ethical production if it could (while the supplier would not chose it on its own). As the headquarter can extract the full expected profits of the match, it seeks to maximize *joint profits* (while the supplier trades off the cost savings only against its own fraction of the expected revenues). Define the technology cutoffs $\beta_{H,k}$ with $k \in \{V, O\}$ as the cutoff headquarter intensities above which joint profits are maximized by ethical production. The supplier only chooses ethical production for $\beta > \beta_S$. We will see below that $\beta_{H,V} < \beta_{H,O} < \beta_S$. Then, the second condition is given by $\beta \in (\beta_{H,O}, \beta_S)$ (condition 2): in this range the headquarter *would like* the supplier to produce ethically (and would then like to choose integration as long as condition 1 is satisfied). But the supplier will implement the unethical technology. Under condition 1 this implies that outsourcing is chosen by the headquarter in order to incentivize the supplier to expand unethical production. Therefore, if there is a non-empty set of headquarter intensities that simultaneously satisfy conditions 1 and 2, the described tension between aspirations and reality emerges as an equilibrium outcome. The following Proposition establishes that this is the case.

Proposition 5 *The technology cutoffs maximizing joint profits satisfy $\beta_{H,V} < \beta_{H,O} < \beta_S$. There is a non-empty set of headquarter intensities that satisfy $\beta \in (\beta_{H,O}, \beta_S) \wedge \beta \in (\beta_e, \beta_u)$. That is, the headquarter would oblige the supplier to produce ethically if it could, but, as it cannot, chooses outsourcing in order to expand unethical production.*
Proof: *See the Appendix.*

3 Firm Choices and Boycotts: a Microfoundation with Private Information

In the baseline model we simply imposed that any deviation from the optimal ethical investment triggers a boycott. We alluded to the intuition that this is necessary in order to keep the unethical practices under cover. This allowed us to focus our analysis on the predictions for the international organization of production. In this section we present an extension of the model that features private information on technology and an advocacy NGO investigating firms. We will take a clear stand on how consumer boycotts emerge, how unethical production affects the risk of facing a boycott as well as the resulting investment and pricing decisions as equilibrium outcomes. We show that the qualitative results of the baseline model and the empirical prediction we derive from it continue to hold in this microfounded extension of the model.

The main intuition behind the microfoundation is that there is an incentive for the supplier to keep unethical production under cover when consumer boycotts can occur in response to it. Maintaining at least some degree of uncertainty about the technology used on the side of consumers or an NGO may reduce the probability of facing a consumer boycott. In fact, acquisition of verifiable information on pollution and working conditions and the link to final consumer brands is a costly and possibly dangerous (and illegal) activity in many countries. A case in point is the Detox campaign by Greenpeace addressing, among other things, the toxic water pollution of the Pearl and Yangtze River Deltas (Greenpeace, 2011) and the Qiantang River (Greenpeace, 2012) in China by local textile and apparel producers. According to Greenpeace, a year-long investigation into production practices and buyer-seller linkages preceded its campaign to push a large number of top labels in the apparel industry to 'detox' their supply chain. Another well-mediatized example of the dangers of investigating working conditions in countries like China is the case of a labor activist being arrested for trying to document poor working conditions in a factory producing shoes for Ivanka Trump's brand in southern China (Bradsher, 2017).

We argue that this strong preference for discretion regarding pollution and working conditions, even backed by national legislation in key countries like China, is an important feature worth modeling explicitly. Our microfoundation therefore grounds on the technology implemented by the supplier being private information of the firm and being costly to verify by a third party.

3.1 Private Information, NGO and Consumer Boycotts

Private information about the type of technology implies that the technology cannot be directly observed from outside the match (while the headquarter and the supplier observe it). Other variables like organization of production, investments, output and prices are observable.

In the baseline model, either all firms in a sector choose the ethical technology or all choose the unethical technology. This is a very stylized pattern that directly stems from the fact that all firms in a sector are identical. In a sector in which all firms implement the unethical technology, mimicking would not make sense, as there are no ethical firms to mimic. We therefore assume that only a fraction κ of suppliers in each sector is able to use the unethical technology. Because of this, in equilibrium there will be at least a fraction $1 - \kappa$ of firms that produce ethically.

In period zero, when the headquarter offers the transfer payment to the supplier and decides the organizational form of the firm, neither party knows whether unethical production will be possible. This is only revealed at the next stage just before investment decisions are taken and the (un)ethical technology choice is made.[23] This assumption implies that the organizational choice of the firm does not contain information on the type of the firm: when it is taken, the headquarter does not know whether the unethical technology will be available in period $t_1(a)$.

In contrast to the baseline model, we now have to be more specific about how a consumer boycott emerges. We assume that there is an NGO that is able to organize such boycotts. Different to e.g. Krautheim and Verdier (2016) or Aldashev and Verdier (2009) the focus of this paper is not on the endogenous emergence of NGO activity, the trade-offs NGOs face, competition among them or the possible types of agents (motivated or not) that run the NGOs. We therefore keep the NGO very stylized.

The objective of the NGO is simply to start boycotts against as many unethical firms as possible. An NGO can trigger a consumer boycott if it can provide sufficient proof that a supplier has implemented an unethical technology. For simplicity, we assume that triggering the boycott is costless for the NGO, while proving the use of the unethical technology is (potentially) costly.[24]

The NGO is sophisticated enough to determine the optimal choices of an ethical firm in a given sector. An unethical firm can choose to pool with ethical firms along all observables (mimicking). In this case, the NGO faces a group of *seemingly ethical firms* that are all identical in terms of observables, but which contains ethical and unethical firms. In this case the NGO has to incur a cost to identify unethical firms and to collect sufficient proof to build a campaign upon. We assume that the NGO has sufficient resources to run costly investigations on a fraction $1-\gamma$ of firms. An unethical firm can also choose to be openly unethical making choices that are inconsistent with ethical production. We assume that this is sufficient proof for the NGO to build a campaign upon.[25] The belief formation of the NGO is discussed in detail below.[26]

We will see below that the investment choices of the headquarter and the supplier are the key variables potentially providing information on the type of the firm to the NGO. This changes the best response

[23] One way to think about this is as follows. Ex-ante the supplier knows that there is some probability κ that it can e.g. bribe government officials to turn a blind eye on toxic waste disposal into a river or on the violation of work safety standards. If this is actually possible in the individual case, only turns out after the match is formed and some investments are made.

[24] This is the simplest possible way to assure that each firm that is identified as unethical also faces a boycott. The alternative would be assuming a sufficiently high (exogenous) NGO budget. This allows us to abstract from the additional complexity of endogenizing the probability of facing a boycott for an identified unethical firm.

[25] The assumption of zero cost of identifying and boycotting an openly unethical firm is again convenient for tractability, but could be relaxed - preserving tractability - if a budget of the NGO high enough to investigate all unethical firms is assumed. Being openly unethical will not be an equilibrium outcome, as this results in zero demand with certainty. This is a very useful benchmark that generates damage by a campaign in the starkest possible way and greatly simplifies the analysis.

[26] One may think that the fact that the NGO interprets the investment levels as containing information of the type of the firm can only work in a context of homogeneous firms. And indeed, when firms differ in productivity (and if this productivity is private information to the firm) different investment levels would be in line with ethical production. One could probably construct a complicated argument on how the NGO forms expectations on the probability of unethical production conditional on observing the investment level and accounting for the underlying productivity distribution. The NGO may then assign a higher probability to controlling firms with 'unlikely' investment levels. There is, however, a very simple alternative way to include firm heterogeneity into the model without raising such concerns. Already Melitz (2003) highlights that heterogeneity in technology (differences in productivity) or in preferences (differences in quality) are isomorphic in his model. For the latter case, it is quite obvious that being part of the utility function of the consumer, quality can hardly be private information of the firm. So conditional on the - observable - quality, the ethical investment level can again be computed. Therefore, an extension of the model to a setting with heterogeneous firms would not be inconsistent with our microfoundation. As argued in Footnote 13 in the introduction, we doubt that the additional insights of such an extension would outweigh the likely costs in terms of tractability.

functions in the non-cooperative investment game. In particular, there will be one single investment profile (the one the NGO expects from an ethical firm) that leads to positive demand in expectation.

These are the extensions and refinements we make to the baseline model in order to microfound the assumption that any deviation from optimal ethical investments triggers a boycott. All other events in the different periods are just like in the baseline model. Transfer payment and organizational choice take place in t_0. In $t_1(a)$, the supplier first observes whether it can use the unethical technology and then chooses its preferred one. Both supplier and headquarter then set investments non-cooperatively to maximize their respective profits in $t_1(b)$. In t_2, intermediates are produced and create the hold-up problem. In t_3, nature decides which of the firms that are not openly unethical are undergoing a costly investigation by the NGO. The NGO spends all its resources and monitors a fraction $1-\gamma$ of firms and starts a boycott against all firms it finds to be unethical. Supplier and headquarter renegotiate the distribution of revenue in period t_4 and in t_5, final goods are produced, sold, and the resulting revenue is distributed to both parties according to the rule established in the bargaining at t_4. We will next discuss the informational content of the firms' choices as well as belief formation of the NGO.

3.2 Setting a 'Signal' Non-Cooperatively?

There are three variables that are observable to the NGO and that potentially contain information on the type of technology implemented. Investments, the quantity produced, and the prices set. The organization of production (outsourcing vs. integration) is decided upon in period zero, which is before nature decides whether the unethical technology is available to the supplier. We argued in Footnote 18 that the produced quantity and the price directly follow from the investment decisions. This implies that the investment stage is decisive for the signaling considerations.

When the investment and pricing decisions of the firm are interpreted by the NGO as containing information on the implemented technology, there is room for strategic signaling when setting investments and prices. This would place us in the context of a signaling game similar to the one in Krautheim and Verdier (2016). The core idea of the signaling literature in economics (Spence 1973, 1974) is that an agent of a given type may take an action that signals its type to a principal. This can be done by taking actions that are only optimal for one type of agent but not for another. The principal can interpret the choice of the agent as an attempt to signal its type. This requires that both parties understand that the other party takes decisions rationally and that an action can be driven by the intention to signal one's type by taking an otherwise suboptimal action.

The obvious difference to our setting is that investments - the decision that contains information about the type of the firm - is taken non-cooperatively. So there is not one agent rationally choosing an investment in order to signal its type: headquarter and supplier cannot coordinate to choose the profit maximizing investment, so they cannot coordinate on an investment in order to signal their type either. This implies that we are not in the context of a signaling game.

While investments cannot be set in a strategic attempt to signal the type of the firm, they are still interpreted by the NGO as potentially containing information on the firm type. In the case of an unethical firm this means that the 'wrong' investment choices can trigger an investigation by the NGO and lead to a boycott. We will see below that this changes the best response function of headquarter and supplier in the non-cooperative investment game.

3.3 NGO Beliefs and Investigations

The only difference between firms in a sector is whether they have the option to implement the unethical technology. The ex-ante probability that a given supplier has this option is given by κ. Here, we are interested in the question what optimal choices of an ethical and unethical firm are conditioning on their type $\theta \in \{e, u\}$. Whether the firms with an option to produce unethically actually decide to do so, is determined at an earlier stage.

When unethical production is profitable in expectation, the NGO knows that a fraction κ of firms are unethical. The non-cooperative investment game results for each firm in an observable investment profile $i(\theta) = \{h(\theta), m(\theta)\}$ with $h(\theta) \geq 0$ and $m(\theta) \geq 0$.[27] In period t_3 the NGO picks an action $s_i \in \{0, 1\}$ which is to initiate an investigation on firms with investment profile i.

The NGO has a belief function $\eta(\theta \mid i)$. Conditional on observing some investment profile i, it assigns a probability of $\eta(\theta \mid i)$ to the firm being of type θ. If $\eta(\theta = u \mid i) = 1$, the NGO immediately starts an investigation.

Proposition 6 *In the extended model,*

(i) ethical firms are indifferent to NGO investigations and therefore set their investments independently of NGO beliefs.

(ii) unethical firms face an NGO investigation with certainty unless they mimic (i.e., set the same investment as) ethical firms. If unethical firms mimic ethical firms, their probability of being investigated is reduced to $1 - \gamma < 1$.

Proof: *In the text.*

The expectations of the NGO follow Bayes' Law implying the following belief function

$$\eta(\theta = e \mid i) = \frac{Pr(i \mid \theta = e)\ Pr(\theta = e)}{Pr(i \mid \theta = e)\ Pr(\theta = e) + Pr(i \mid \theta = u)\ Pr(\theta = u)}. \tag{22}$$

Note that ethical firms are indifferent to being investigated: they always get full demand in period t_5, as they never face a boycott. Denote by $\tilde{i}$ the investment profile of an ethical firm resulting from the non-cooperative investment game. An ethical firm would never adjust $\tilde{i}$ to accord with an arbitrary belief of the NGO, as this only affects the probability of being investigated, which has no effect on the firm.

We therefore have $Pr(\tilde{i} \mid \theta = e) = 1$ and $Pr(\bar{i} \mid \theta = e) = 0$ for any $\bar{i} \neq \tilde{i}$.[28] Therefore, $\tilde{i}$ is the only investment profile for which the NGO assigns a positive probability to ethical production: $\eta(\theta = e \mid \tilde{i}) > 0$ and $\eta(\theta = e \mid \bar{i}) = 0$ for any $\bar{i} \neq \tilde{i}$. Any other investment profile triggers an immediate investigation by the NGO.

[27] For ease of exposition we suppress the organizational subscript k and the variety index ω where possible. It is well understood that the strategies are chosen and decision are made conditional on outsourcing or vertical integration chosen by the headquarter at an earlier point in the game.

[28] In a signaling setup, one would have to further investigate the question if ethical firms would want to deviate from $\tilde{i}$, choosing an investment profile that is unprofitable to mimic for unethical firms. As outlined above, in our model investments are not contractible and are set non-cooperatively. Therefore, investments cannot be used to signal the type of the firm to the NGO.

The NGO can compute if in a given sector firms have an incentive to be unethical. When unethical firms in that sector pool with ethical firms by setting $\tilde{i}$, they form a group of *seemingly ethical firms* for which investigation is costly for the NGO. As in this case $\eta(\theta = e \mid \tilde{i}) < 1$, the NGO trivially maximizes its objective of starting a boycott against the largest possible number of unethical firms by spending its whole budget on investigations of firms in the seemingly ethical group (and then start costless boycotts against all identified unethical firms). By assumption its exogenous budget is sufficient to investigate a fraction $1 - \gamma$ of these firms.

3.4 Non-Cooperative Investments with Degenerate Demand

We have seen above that unethical firms can only generate positive demand (in expectation) by investing $\tilde{i}$. For this investment the firm faces full demand if it arrives at stage t_5 without a boycott.

Lemma 1 *The equilibrium investment profile $\tilde{i}$ of an ethical firm is characterized by the same expressions, i.e. equations* (9) *and* (10)*, as the equilibrium profile i^* in the baseline model.*
Proof: *This directly follows from the fact that the optimal choices of the headquarter and the supplier in a match that only has the ethical technology available (or in a sector where all firms endogenously choose ethical production), is unaffected by any element of the microfoundation.*

It remains to be shown that $\tilde{i} = i^*$ is the equilibrium outcome of the non-cooperative investment game also for an unethical firm. Clearly, it is a Nash equilibrium of the investment game if it yields positive profits in expectation, as any deviation from it would lead to zero demand. As in the Antràs (2003) model, zero-zero is a Nash equilibrium that is ruled out by the Pareto dominance assumption.

Consider the case of an unethical firm (i.e. the decision to use the unethical technology has already been taken). The right-hand side graph in Figure 2 illustrates the best responses of the investment game in this case. The best response to any investment level other than $i_k^* = \{h_k^e, m_k^e\}$, with $k \in \{V, O\}$ is zero for both parties, as any deviation from i_k^* leads to an investigation by the NGO resulting in a boycott with zero demand. No party would ever find it optimal to choose an investment that is not on its best response function, as it would be strictly dominated by playing the best response. We can therefore state the following proposition.

Proposition 7 *In the extended model, unethical firms mimic ethical firms, i.e. the equilibrium investment profile of an unethical firm is identical to the equilibrium investment profile of an ethical firm.*
Proof: *In the text.*

Using the results of this Section, we show in the Online Appendix that the microfounded version of our model produces the same qualitative results as the baseline model.

4 Implementation of the Empirical Test

In this section, we provide empirical support for the main prediction of our theory. To do so we follow the established empirical literature, especially Nunn and Trefler (2013), Antràs and Chor (2013) and Antràs and Yeaple (2014), as our main references. We use U.S. Census Bureau data on intrafirm trade for the years 2007 to 2014 and employ the standard measure of vertical integration at the industry level: the share of US intrafirm imports in total U.S. imports.

One reason for using industry-level data is that the very few firm-level datasets that contain information on organizational decisions are not publicly available.[29] Second, we need information on the incentives to produce unethically, which, to our knowledge, are not available in these data.

We test the prediction of the model summarized in Proposition 4: a strong cost advantage of unethical production in an industry increases the prevalence of international outsourcing relative to vertical integration. The actual cost advantage of unethical production has an industry and a country component. We propose a measure for an industry's potential for cost savings when operating under looser environmental regulation. These *potential* cost savings only translate into *actual* cost savings if production takes place in a jurisdiction with weaker regulation than the U.S.We therefore also provide an appropriate measure for the strength of environmental regulation in different countries. We discuss these variables in detail below.

4.1 Intrafirm Import Share

Data on intrafirm trade at a detailed country-industry level come from the Related Party Trade Database administered by the U.S. Census Bureau.[30] We use information on U.S. imports in manufacturing from all over the world at the NAICS 6-digit level for the years 2007 to 2014. We convert the data to IO2007 industries from the BEA's input-output tables.

Crucially, the trade flows are distinguished by the relationship between the entities who trade them. A trade flow is marked as taking place between two related parties when the importer holds at least a 6% equity stake in the exporter and as unrelated trade otherwise. We construct our dependent variable, the intrafirm import share, as the value of related party imports over the sum of the value of related and unrelated party imports for each IO2007 industry-country-year.[31] Our regression sample includes 231 manufacturing industries in the least data-demanding specification. In our main specifications, we have around 215 industries.

4.2 The Unethical Environmental Cost Advantage

The key parameter in our model is the *unethical cost advantage* $1-\mu$. μ measures the ratio of unethical to ethical marginal cost of the manufacturing input in the model, and we stress that the input m in the model stands for everything the supplier contributes to the production process. The marginal cost c_m therefore captures not only payments to workers, but also other expenditures by the supplier, such as provisions for workplace safety and the cost of compliance with local environmental regulation. For reasons of data availability, we focus our analysis on the *environmental* incentives of unethical production.

[29] Antràs (2013) names four such datasets. A comprehensive Japanese manufacturing survey used by Tomiura (2007), administrative French data used by Corcos, Irac, Mion, and Verdier (2013) and Defever and Toubal (2013), a Spanish survey used by Kohler and Smolka (2014) and the WorldBase database by Dun & Bradstreet, used in Alfaro, Conconi, Fadinger, and Newman (2016).

[30] The data are available online from http://sasweb.ssd.census.gov/relatedparty/.

[31] A third category, unreported trade, captures import flows that are not marked as either type of trade. The share of unreported trade flows in total imports is usually negligible. Antràs and Chor (2013) provide a more detailed discussion of the distribution of unreported trade across industries and source countries.

4.2.1 The Industry Dimension

An industry producing large volumes of toxic waste, carbon dioxide emissions and which is intensive in the use of natural resources arguably benefits more from unethical production. To measure this environmental unethical cost savings *potential* of an industry we draw on data from the Annual Survey of Manufactures (ASM) provided by the US Census Bureau. Starting from 2007, the survey records the industry-level expenditure on water, sewer, and refuse removal, as well as other non-electric utility payments including the cost of hazardous waste removal. We use this waste removal expenditure as a proxy for the money amount an industry would save if production took place in an unregulated environment.

One advantage of this measure is that, according to the survey manual, it excludes payments for machinery, equipment, and electric utility.[32] This makes us confident that we capture only those costs that are directly related to the removal of hazardous materials and other waste and that more capital-intensive industries are not mechanically more intensive in waste removal costs.[33]

We construct our variable of an industry's environmental cost savings potential (ECSP) as the log of an industry's expenditure on waste removal relative to its payroll, total cost, or total sales, respectively. We will explain in Section 4.4 why the normalization by total costs and total sales are our preferred specifications. We provide the results for normalization with payroll for direct comparison to the literature and show that our results are not driven by the change in normalization.

In Figure 5 we provide evidence of the variation in our proxy in a histogram of the ECSP calculated as spending on waste removal relative to industry payroll (left panel) and a proxy of total cost of the industry (right panel) across industries and years.[34] The distribution is very right-skewed in both cases and in the bulk of industries spending on hazardous waste removal makes up between 0% and 10% of payroll or between 0% and 1% of total cost.[35]

4.2.2 The Country Dimension

The extent to which the *potential* cost savings translate into *actual* savings depends crucially on the strictness of regulation in the source country. Only if regulation there is more lenient than in the US, can (some of) the potential cost savings be realized.

To measure the country dimension of the unethical environmental cost advantage, we employ the Environmental Policy Stringency Index (EPSI) computed by the OECD for 26 member countries (excluding the U.S.) and the six non-member countries Brazil, China, India, Indonesia, Russia, and South Africa for the years 2007 to 2012. The index combines information on 14 environmental policy instruments that are mainly related to air and climate pollution and is suitable for comparisons across countries. According to

[32] The survey manual contains detailed instructions about the forms to be filled out by sampled establishments. The manual for the survey year 2015 is available from https://bhs.econ.census.gov/bhs/cosasm/ASMInstructions.pdf. The instruction pertaining to our variable can be found on p.17 of that manual.

[33] Another advantage is that it appears plausible to consider expenditure on waste removal a lower bound for the unethical environmental cost advantage of an industry for two reasons. First, because our measure excludes salaries of employees whose work includes waste removal or treatment. Second, because we are using data from a technologically advanced country, it is likely that the implemented technology in the US is less environmentally intensive than in most other countries. It is therefore likely that production in many other countries takes place with more environmentally intensive technologies implying that the true potential cost savings are likely to be higher than measured by our variable.

[34] The construction of the proxy for total cost is described in Section 4.6.2.

[35] In the Online Appendix we document that our measures generate rankings of industries that are arguably in line with common preconceptions about environmentally 'dirty' industries.

the OECD's definition, a policy is more stringent if it puts a higher explicit or implicit price on pollution or environmentally harmful behavior. An index value of 0 is the lowest stringency possible, while an index value of 6 denotes the highest stringency. The maximum value the index attains in our sample is 4.41 for Denmark in 2009. The lowest value is .375 for Brazil in 2011.

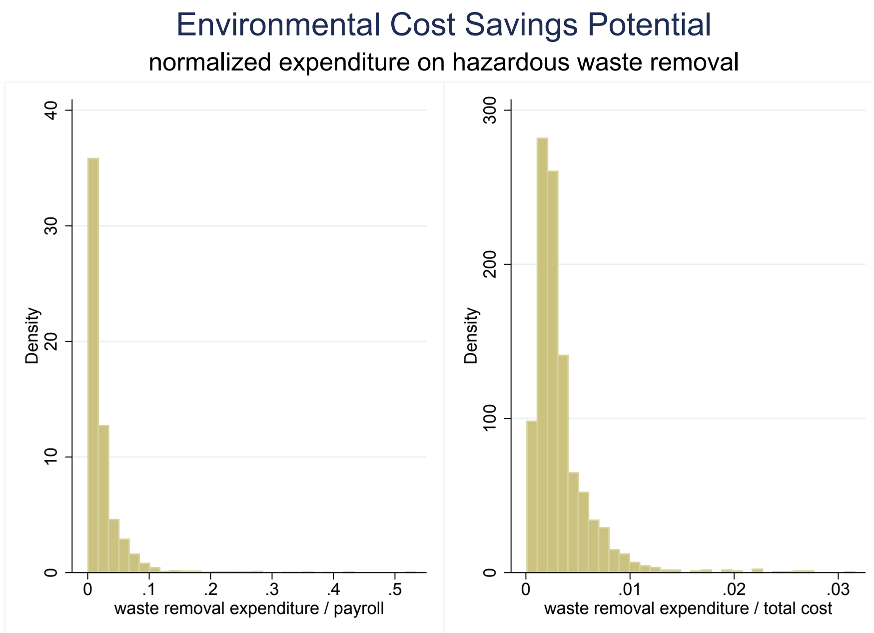

Figure 5: Variation in environmental cost savings potential across industries and over time.

4.3 Control Variables

In addition, we use various control variables that have been identified in the literature as determinants of intrafirm trade or have been used for robustness checks therein. See Nunn and Trefler (2013), Antràs and Chor (2013) and Antràs and Yeaple (2014). In particular, we control for the logs of capital intensity, R&D intensity and high-skill intensity. We take the data on physical capital expenditure and the share of non-production worker wages from the ASM. R&D intensity is defined as R&D expenditure relative to sales and is calculated from Compustat data on U.S. firms. In addition, we control for material intensity (normalized expenditure on materials, from ASM) and the within-industry size dispersion computed by Nunn and Trefler (2008). We follow the literature and disaggregate capital into its components, which arguably differ in relationship-specificity, to obtain a cleaner proxy for headquarter intensity.[36]

4.4 Intensities

We normalize our explanatory variables by industry payroll to make them consistent with the construction of the proxies for headquarter intensity in the literature.[37] Regressions using this definition will provide an easy point of reference to compare our results with those from the preceding literature.

[36]Further details on the sources of these variables and their construction are delegated to the Data Appendix.
[37]An exception is R&D intensity, which is normalized by total sales in the literature, which we also follow for comparability.

In our preferred specifications, however, we construct all intensities (except for R&D) as the log of the respective expenditure relative to total industry cost. We assess the robustness of our results by normalizing with total sales as well. While total industry sales can be taken directly from the data, we must construct a proxy for total cost, for which we sum payroll, cost of materials, total capital expenditure, total rental payments and an aggregate term for all other expenditures from the ASM. We prefer these definitions because we believe they capture more directly the relative importance of a particular type of cost for the overall production process. As explained in Section 2.1.2, we interpret the factors of production in the model as aggregate inputs each party brings into the production relationship. Different types of costs play more or less significant roles in these aggregates. For example, firms typically spend on R&D, invest in physical capital and hire labor at the same time. They also incur other types of costs, including expenditure on the removal of (hazardous) waste or investments in workplace safety or costs of acquiring inputs and intermediate products. Some of these costs tend to be incurred by the headquarter, others by a supplier (integrated or independent). In our view, all these different types of costs should be accounted for when factor intensities are computed as the question we seek to address is: Do industries outsource *more* in low regulation countries, when they can potentially save a *larger fraction* of *total cost* by producing unethically?

An additional argument for using broader measures of factor intensities is the fact that the share of capital expenditure in total cost and the share of the wage bill (payroll) in total cost are significantly correlated with a positive coefficient of 0.1345 in our data.[38] This casts some doubt on the consistency of computing capital intensity - the key variable in the existing literature - by normalizing by payroll in our sample.

4.5 Empirical Specification

We estimate variants of the following regression equation.

$$intrafirm_{ijt} = \eta_0 + \eta_1 \; ECSP_{jt} + \eta_2 \left(ECSP_{jt} \times EPSI_{it}\right) + \rho X_{jt} + \zeta_{it} + \epsilon_{ijt}. \tag{23}$$

$intrafirm_{ijt}$ is the share of related party imports in total imports by the US from country i in industry j in year t. $ECSP_{jt}$ is our proxy for the part of the *unethical environmental cost advantage* varying across industries j and over time t. $EPSI_{it}$ proxies for the part of the *unethical environmental cost advantage* that varies across source countries i and time t. X_{jt} contains the established determinants of intrafirm trade and the other control variables mentioned above. ζ_{it} is a set of country-year fixed effects to control for everything that is specific to a country in a given year. The fixed effects therefore control for the level effect of the $EPSI_{it}$. They also control for the endogenous choice of a sourcing location to the extent that this is driven by country- or country-year-specific factors, such as geography, corporate tax rates or cultural linkages. We want to take out this variation to be able to make statements about the tendency to outsource production conditional on the chosen source country. In all our regressions, we cluster standard

[38] The correlation coefficient is .1687 when using total sales in the denominator. This correlation is puzzling when one has a Cobb-Douglas production function in mind with labor and capital as inputs. In the data, a very large portion of an industry's expenditure is allocated to intermediate inputs. When we correlate the sum of payroll and material input expenditure relative to total cost with the share of capital expenditure in total cost, the correlation coefficient is highly significant at -0.5677.

errors at the IO2007-industry level as this is the level of variation of our main explanatory variables and industry characteristics are highly auto-correlated over time.

Our data on intrafirm imports cover 230 countries and territories. But our measure of the level of regulation, EPSI, is limited to 26 OECD countries (excluding the US) plus the six non-member countries listed above. We therefore run the specification in equation (23) in two versions.

In the first specification we only include $ECSP_{jt}$ but not the interaction effect. This allows us to make use of the full sample. In this case the prediction of the model holds under the premise that most of the 230 countries and territories have more lenient regulation and enforcement (capacity) than the U.S. Within the set of countries for which EPSI data are available the US takes a middle position. Arguably, many, if not most, of the 198 countries and territories for which EPSI is *not* available (the remaining non-OECD countries plus OECD members Chile, Estonia, Iceland, Israel, Latvia, Luxembourg, Mexico, and New Zealand) should indeed be expected to have more lenient regulation and enforcement (capacity) than the US. The presence of some countries with similar or higher levels of regulation should bias the results against our hypothesis, so it is save to keep them in the sample.[39] We therefore expect $\eta_1 < 0$: industries with a higher potential cost advantage should have a lower share of intrafirm trade.

In the second specification, we add the interaction of the cost savings potential $ECSP_{jt}$ and $EPSI_{it}$, the OECD Environmental Stringency Index. Due to the limited coverage of the $EPSI_{it}$ we have a strongly reduced sample size in this specification. On the other hand, the interaction effect allows for more flexibility to analyze the differential impact a given level of $ECSP_{jt}$ has across varying regulatory environments. The tendency to outsource production in industries with a given $ECSP_{jt}$ should be stronger when the goods are sourced from countries with more lenient environmental policies. In the second specification, we therefore expect $\eta_2 > 0$ and continue to expect $\eta_1 < 0$.

4.6 Empirical Results

In this subsection we present our estimation results. In our preferred specification, we normalize the explanatory variable with total cost. We then show that the results we find also hold qualitatively when we normalize with total sales and payroll. First, however, we show that the well-established results in this literature also hold in our data.

4.6.1 Previous Literature

Replication of earlier results provides a useful benchmark for our empirical work as we use data from the same sources but for the years 2007 to 2014.[40] Intensities are constructed relative to industry payroll. In the case of R&D intensity, we follow Antràs and Yeaple (2014) and add 0.001 to the ratio of R&D expenditure over sales before taking the natural log in order to avoid throwing away the zeros.[41]

[39] We have experimented with leaving out countries with a stricter EPSI value than the US based on the OECD data. As expected, this changes significance levels and coefficient mildly in favor of our hypothesis.

[40] Nunn and Trefler (2013) use data for the year 2005 only, Antràs and Chor (2013) use data from 2000 to 2010. Antràs and Yeaple (2014) use data from 2000 to 2011 from the intrafirm trade data and shorter subsets of this time span for the industry controls.

[41] We recognize that this way of handling zeros is not innocuous but follow the literature to ensure comparability. We have experimented with other values, such as adding 0.00001 as Nunn and Trefler (2013) do, and this does not change our results qualitatively.

Column 7 of Table 1 reports results of a regression specification as in the previous literature, including the established decomposition of capital into its components and normalization with payroll. Other machinery is arguably the most relationship-specific of the four capital components and is strongly associated with more intrafirm trade. R&D intensity and dispersion are also highly significant and positively associated with intrafirm trade. These results are consistent with prior evidence on the determinants of intrafirm trade. In columns 1 and 4 we rerun the established specification using total cost and total sales, respectively, as normalization variable. The results are quite similar, both quantitatively and qualitatively.

4.6.2 Core Findings

In column 2 of Table 1 we add our measure of environmental cost saving potential (ECSP) to our preferred specification with the total cost normalization. We indeed find that a larger ECSP is associated with less intrafirm trade on average and is significant at the 10% level. The other coefficients do not change much compared to column 1 and continue to have the right signs. Industries with a higher ECSP seem to be more likely to outsource production. The number in brackets reports the standardized beta coefficient associated with the respective coefficient. When the log of the ECSP increases by one standard deviation, the intrafirm trade share decreases by 4.5% of a standard deviation on average.

In column 3 we add the interaction term of the ECSP with the index of environmental policy stringency (EPSI). As expected, we find the interaction effect to be positive and significant at the 5%-level. The level effect of the ECSP almost doubles in absolute magnitude and is negative and significant at the 1% level. The interaction effect uncovers a strong cross-country pattern of heterogeneity in the effect of the ECSP. This underscores the empirical importance of both the industry-specific and the country-specific components of the parameter $1 - \mu$.

Due to to the limited coverage of the EPSI our sample drops to roughly one fourth of its former size as we have to exclude the many non-OECD countries (except the six emerging economies mentioned above) for which we do not have data. In the Data Appendix we provide additional tables which show that the level effect of the ECSP is also negative when we remove the interaction effect and hold the (small) sample size constant. In many cases the level effect is not significant when the sample size is reduced, indicating that it is indeed countries outside the realm of developed OECD countries driving our results.

To analyze the cross-country dimension further, we report marginal effects of the ECSP at various percentiles of the distribution of the EPSI. In Table 2, columns 2 and 3 show the marginal effect and the corresponding p-value for the total cost specification from Table 1. There is sizable variation in the marginal effect. The coefficients are significant at the 1% level up to and including the first decile. The four countries in the first decile are Brazil, China, Indonesia, and South Africa. They turn insignificant by conventional levels at the fourth decile. The marginal effect continues to fall until it reaches a value of almost zero at the ninth decile of our sample. Table 2 clearly shows that the effect of the ECSP on intrafirm trade is driven by the countries with the lowest environmental regulation. This supports our theoretical setting in which the possibility of environmentally unethical production arises due to differences in regulation across countries.

Dividing our explanatory variables by total sales in columns 4 to 6 of Table 1, it becomes clear that our result is not driven by the normalization variable we use. The coefficient of the ECSP is weakly

Table 1: The Effect of Unethical Environmental Cost Advantage on Intrafirm Trade

Dependent Variable: Intrafirm Import Share									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intensity Definition:	Total Cost	Total Cost	Total Cost	Total Sales	Total Sales	Total Sales	Payroll	Payroll	Payroll
log ECSP		-0.0225*	-0.0401***		-0.0200*	-0.0387***		-0.0238**	-0.0270*
		(0.0122)	(0.0143)		(0.0120)	(0.0143)		(0.0115)	(0.0139)
		[-0.0446]	[-0.0897]		[-0.0388]	[-0.0843]		[-0.0623]	[-0.0783]
log ECSP			0.00893**			0.00917**			0.00174
X EPSI			(0.00428)			(0.00435)			(0.00410)
			[0.176]			[0.187]			[0.0263]
log other machinery intensity	0.0299***	0.0401***	0.0562***	0.0306***	0.0399***	0.0503***	0.0276***	0.0405***	0.0494***
	(0.0110)	(0.0108)	(0.0140)	(0.00974)	(0.0101)	(0.0145)	(0.00948)	(0.00993)	(0.0139)
	[0.0490]	[0.0657]	[0.104]	[0.0494]	[0.0643]	[0.0909]	[0.0629]	[0.0922]	[0.124]
log skill intensity	0.0402*	0.0383*	0.0586***	0.0350***	0.0354***	0.0375**	0.0497*	0.0529*	0.0498
	(0.0221)	(0.0213)	(0.0208)	(0.0119)	(0.0117)	(0.0151)	(0.0282)	(0.0276)	(0.0372)
	[0.0837]	[0.0797]	[0.133]	[0.0739]	[0.0746]	[0.0864]	[0.0445]	[0.0473]	[0.0495]
log R&D intensity	0.0221***	0.0208***	0.0267***	0.0224***	0.0211***	0.0279***	0.0214***	0.0197***	0.0269***
	(0.00390)	(0.00402)	(0.00482)	(0.00382)	(0.00394)	(0.00495)	(0.00450)	(0.00451)	(0.00543)
	[0.0960]	[0.0903]	[0.130]	[0.0970]	[0.0914]	[0.135]	[0.0928]	[0.0855]	[0.131]
log materials intensity	0.0747	0.0595	0.131**	0.0554**	0.0498**	0.0455	-0.00568	0.00238	-0.0101
	(0.0637)	(0.0629)	(0.0585)	(0.0224)	(0.0246)	(0.0328)	(0.0115)	(0.0114)	(0.0117)
	[0.0405]	[0.0322]	[0.0790]	[0.0391]	[0.0352]	[0.0366]	[-0.0118]	[0.00495]	[-0.0229]
dispersion	0.0840***	0.0790***	0.0861***	0.0836***	0.0792***	0.0871***	0.0821***	0.0766***	0.0849***
	(0.0138)	(0.0134)	(0.0139)	(0.0139)	(0.0135)	(0.0146)	(0.0137)	(0.0131)	(0.0147)
	[0.0966]	[0.0911]	[0.114]	[0.0962]	[0.0913]	[0.115]	[0.0944]	[0.0883]	[0.112]
log building intensity	-0.0108*	-0.00918	-0.0118	-0.00844	-0.00735	-0.0103	-0.0138**	-0.0116*	-0.0149*
	(0.00643)	(0.00647)	(0.00780)	(0.00638)	(0.00634)	(0.00742)	(0.00608)	(0.00627)	(0.00780)
	[-0.0227]	[-0.0194]	[-0.0283]	[-0.0174]	[-0.0152]	[-0.0241]	[-0.0347]	[-0.0291]	[-0.0415]
log auto intensity	-0.0116**	-0.0118***	-0.0181***	-0.0133***	-0.0133***	-0.0211***	-0.0106**	-0.0107**	-0.0181***
	(0.00457)	(0.00436)	(0.00596)	(0.00459)	(0.00444)	(0.00615)	(0.00470)	(0.00443)	(0.00627)
	[-0.0351]	[-0.0357]	[-0.0569]	[-0.0410]	[-0.0406]	[-0.0677]	[-0.0304]	[-0.0304]	[-0.0537]
log computer intensity	-0.00912	-0.0121	0.000706	-0.00693	-0.0101	0.00337	-0.0117	-0.0148*	-0.00243
	(0.00765)	(0.00779)	(0.0106)	(0.00711)	(0.00751)	(0.0108)	(0.00769)	(0.00766)	(0.0109)
	[-0.0222]	[-0.0294]	[0.00181]	[-0.0167]	[-0.0243]	[0.00859]	[-0.0224]	[-0.0284]	[-0.00486]
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IO2007 industry clusters	212	212	212	212	212	212	212	212	212
Observations	130,985	130,402	35,434	130,985	130,402	35,434	130,985	130,402	35,434
R-squared	0.179	0.181	0.169	0.181	0.182	0.169	0.179	0.181	0.167

Note: Estimation by OLS with standard errors clustered at the industry level reported in parentheses. Standardized beta coefficients reported in brackets. ***, **, and * denote significance the 1%, 5%, and 10% level, respectively. log ECSP is the log of expenditure on waste and hazardous materials removal over payroll, total cost or total sales.

significant and negative in column 5, and becomes larger in absolute terms and highly significant when we add the interaction term, which is positive and significant at the 5%-level here as well. Turning to the marginal effects in columns 4 and 5 of Table 2, it is evident that the pattern is qualitatively and quantitatively almost identical to the one from the total cost specification.

In Columns 7 to 9 of Table 1, we test our prediction using the established payroll definition of intensities. When included by itself in column 8, the effect of the ECSP is negative with roughly the same magnitude as the coefficients from columns 2 and 5. It is even significant at the 5%-level. When we add the interaction effect in column 9, the pattern holds qualitatively, with a negative level effect and a positive interaction term. However, significance levels are lower than in the other specifications. This result is mirrored in columns 6 and 7 of Table 2. The magnitude of the marginal effect changes only very little over the distribution of the EPSI while significance levels range from 5% below the median and a 10%-level of significance up to the ninth decile.

Table 2: Marginal Effects of the ECSP

Intensity Definition	Total Cost		Total Sales		Payroll	
EPSI percentile	Marginal Effect	p-value	Marginal Effect	p-value	Marginal Effect	p-value
5	-0.036	0.007	-0.034	0.009	-0.026	0.043
10	-0.035	0.008	-0.033	0.010	-0.026	0.041
20	-0.027	0.023	-0.025	0.031	-0.024	0.035
30	-0.022	0.060	-0.020	0.078	-0.023	0.039
40	-0.020	0.102	-0.018	0.132	-0.023	0.045
50	-0.017	0.153	-0.015	0.196	-0.023	0.052
60	-0.016	0.204	-0.014	0.259	-0.022	0.060
70	-0.014	0.275	-0.012	0.345	-0.022	0.070
80	-0.012	0.375	-0.009	0.461	-0.021	0.085
90	-0.010	0.448	-0.008	0.544	-0.021	0.097

Note: Marginal effects of log ECSP at deciles of the EPSI are calculated from the regressions in Table 1, columns 3, 6, and 9, respectively.

4.6.3 Robustness

We conduct various checks to assess the robustness of the effect we find. In particular, we add a measure of downstreamness and its interaction with the elasticity of substitution as in Antràs and Chor (2013) and include further controls used in that paper as robustness checks. We report the results in the Online Appendix.

5 Conclusion

In this paper we developed a model of the international organization of production with international regulatory differences, unethical cost savings and consumer boycotts. We have shown that a high supplier intensity of the production process favors the implementation of the unethical technology as well as international outsourcing, while headquarter-intensive sectors tend to choose integration and ethical production. The headquarter has no instrument to affect the supplier's technology choice. The implementation of the unethical technology by the supplier, however, feeds back on the headquarter's choice of the boundaries of the firm. When the headquarter anticipates unethical production by the supplier, it is more inclined to keep the supplier at arm's length. This new *unethical outsourcing incentive* therefore creates a link between unethical production and outsourcing from within the logic of the property rights theory of the firm: outsourcing increases the optimal investment of the supplier and thereby increases the cost savings of unethical production. We also show that it is possible that the headquarter would prefer ethical production (if technology was contractible) but incentivizes an expansion of unethical production as an optimal response to contract incompleteness. To focus on the implications of our model for the international organization of production, in the baseline model, we imposed that any deviation from investments, quantities or prices of an ethical firm immediately triggers a consumer boycott. We also analyzed a fully micro-founded version of the model where the link between a deviation from the ethical observables and a boycott emerges from asymmetric information, credence goods and an NGO monitoring suppliers and potentially starting boycotts. We found that all results from the baseline model hold qualitatively. Using U.S. Census Bureau data, we have provided evidence that, as predicted by the model, the share of U.S. intrafirm imports is higher in sectors with a strong unethical cost advantage.

Also in line with the theory, this effect is strongest in countries with a low level of regulatory stringency.

References

Aichele, R., and G. Felbermayr (2015): "Kyoto and Carbon Leakage: An Empirical Analysis of the Carbon Content of Bilateral Trade," *Review of Economics and Statistics*, 97(1), 104–115.

Aldashev, G., M. Limardi, and T. Verdier (2015): "Watchdogs of the Invisible Hand: NGO Monitoring and Industry Equilibrium," *Journal of Development Economics*, 116, 28–42.

Aldashev, G., and T. Verdier (2009): "When NGOs Go Global: Competition on International Markets for Development Donations," *Journal of International Economics*, 79(2), 198–210.

Alfaro, L., P. Conconi, H. Fadinger, and A. F. Newman (2016): "Do Prices Determine Vertical Integration," *Review of Economic Studies*, 83(3), 855–888.

Antràs, P. (2003): "Firms, Contracts, and Trade Structure," *The Quarterly Journal of Economics*, 118(4), 1375–1418.

——— (2013): "Grossman-Hart (1986) Goes Global: Incomplete Contracts, Property Rights, and the International Organization of Production," *The Journal of Law, Economics, and Organization*, 30(Supplement 1), i118–i175.

Antràs, P., and D. Chor (2013): "Organizing the Global Value Chain," *Econometrica*, 81(6), 2127–2204.

Antràs, P., and E. Helpman (2004): "Global Sourcing," *Journal of Political Economy*, 112(3), 552–580.

——— (2008): "Contractual Frictions and Global Sourcing," in *The Organization of Firms in a Global Economy*, ed. by E. Helpman, D. Marin, and T. Verdier, pp. 9–54. Harvard University Press.

Antràs, P., and S. R. Yeaple (2014): "Multinational Firms and the Structure of International Trade," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, pp. 55–130. Elsevier.

Baron, D. P. (2001): "Private Politics, Corporate Social Responsibility, and Integrated Strategy," *Journal of Economics & Management Strategy*, 10(1), 7–45.

——— (2003): "Private Politics," *Journal of Economics & Management Strategy*, 12(1), 31–66.

——— (2010): "Morally-Motivated Self-Regulation," *American Economic Review*, 100(4), 1299–1329.

——— (2011): "Credence Attributes, Voluntary Organizations, and Social Pressure," *Journal of Public Economics*, 95(11-12), 1331–1338.

——— (2012): *Business and Its Environment*. Pearson, seventh edn.

——— (2016): "Self-Regulation and the Market for Activism," *Journal of Economics & Management Strategy*, 25(3), 584–607.

Baron, D. P., and D. Diermeier (2007): "Strategic Activism and Nonmarket Strategy," *Journal of Economics & Management Strategy*, 16(3), 599–634.

Basu, K., and Z. Tzannatos (2003): "The Global Child Labor Problem: What Do We Know and What Can We Do?," *World Bank Economic Review*, 17(2), 147–173.

BERNARD, A. B., AND T. C. FORT (2015): "Factoryless Goods Producing Firms," *American Economic Review*, 105(5), 518–523.

BERNARD, A. B., J. B. JENSEN, S. REDDING, AND P. K. SCHOTT (2013): "Intra-Firm Trade and Product Contractibility," *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-12*.

BESLEY, T., AND M. GHATAK (2007): "Retailing Public Goods: the Economics of Corporate Social Responsibility," *Journal of Public Economics*, 91(9), 1645–1663.

BIRD, Y. Z., J. L. SHORT, AND M. W. TOFFEL (2017): "Organizational Structures and the Improvement of Working Conditions in Global Supply Chains: Legalization, Participation, and Economic Incentives," *Harvard Business School Working Paper*, 18-003.

BLOOMBERG (2018): "Apple Supplier Workers Describe Noxious Hazards at China Factory," `https://www.bloomberg.com/news/articles/2018-01-16/workers-at-apple-supplier-catcher-describe-harsh-conditions`, accessed February 14, 2018.

BRADSHER, K. (2017): "Detained Over Ivanka Trump Factory Inspection, China Labor Activist Speaks," `https://www.nytimes.com/2017/07/10/business/china-labor-ivanka-trump-shoes.html`, accessed February 20, 2018.

BRODA, C., AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(2), 541–585.

CARLUCCIO, J., AND M. BAS (2015): "The Impact of Worker Bargaining Power on the Organization of Global Firms," *Journal of International Economics*, 96(1), 162–181.

CHINA LABOR WATCH (2012): "Samsung Factory Exploiting Child Labor," `http://www.chinalaborwatch.org/upfile/2012_8_8/samsung8-27.pdf`, accessed February 14, 2018.

——— (2015): "The Other Side of Fairy Tales. An Investigation of Labor Conditions at Five Chinese Toy Factories," `http://www.chinalaborwatch.org/upfile/2015_11_20/2015.11.20%20The%20Other%20Side%20of%20Fairy%20Tales%20-%20EN%20final.small.pdf`, accessed February 14, 2018.

——— (2018): "Apple's Failed CSR Audit. A Report on Catcher Technology Polluting the Environment and Harming the Health of Workers," `http://www.chinalaborwatch.org/upfile/2018_01_12/20180116-1.pdf`, accessed February 14, 2018.

CLEAN CLOTHES CAMPAIGN (2017): "Labour Without Liberty. Female Migrant Workers in Bangalore's Garment Industry," `https://cleanclothes.org/news/2018/01/resources/publications/labour-without-liberty-2013-female-migrant-workers-in-bangalores-garment-industry-full-version`, accessed February 14, 2018.

CONE (2013): "2013 Cone Communications Social Impact Study," Research Report, Boston.

COPELAND, B. R., AND M. S. TAYLOR (1994): "North-South Trade and the Environment," *Quarterly Journal of Economics*, 109(3), 755–787.

——— (2004): "Trade, Growth, and the Environment," *Journal of Economic Literature*, 42(1), 7–71.

CORCOS, G., D. M. IRAC, G. MION, AND T. VERDIER (2013): "The Determinants of Intrafirm Trade: Evidence from French Firms," *The Review of Economics and Statistics*, 95(3), 825–838.

Costinot, A., L. Oldenski, and J. Rauch (2011): "Adaptation and the Boundary of Multinational Firms," *Review of Economics and Statistics*, 93(1), 298–308.

Defever, F., and F. Toubal (2013): "Productivity, Relationship-Specific Inputs and the Sourcing Modes of Multinationals," *Journal of Economic Behavior and Organization*, 94, 245–357.

Distelhorst, G., J. Hainmueller, and R. M. Locke (2017): "Does Lean Improve Labor Standards? Management and Social Performance in the Nike Supply Chain," *Management Science*, 63(3), 707–728.

Distelhorst, G., and R. M. Locke (forthcoming): "Does Compliance Pay? Social Standards and Firm-level Trade," *American Journal of Political Science*.

Eckel, C., and H. Egger (2009): "Wage Bargaining and Multinational Firms," *Journal of International Economics*, 77(2), 206–214.

Egels-Zandén, N. (2007): "Suppliers' Compliance with MNCs' Codes of Conduct: Behind the Scenes at Chinese Toy Suppliers," *Journal of Business Ethics*, 75, 45–62.

Egorov, G., and B. Harstad (2017): "Private Politics and Public Regulation," *Review of Economic Studies*, 84(4), 1652–1682.

Feddersen, T. J., and T. W. Gilligan (2001): "Saints and Markets: Activists and the Supply of Credence Goods," *Journal of Economics and Management Strategy*, 10(1), 149–171.

Flammer, C. (2013): "Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors," *Academy of Management Journal*, 56(3), 758–781.

Gereffi, G., and F. Mayer (2006): "Globalization and the Demand for Governance," in *The New Offshoring of Jobs and Global Development*, pp. 39–65. International Institute for Labour Studies.

Greenpeace (2011): "Dirty Laundry. Unravelling the Corporate Connections to Toxic Water Pollution in China," `http://www.greenpeace.org/eastasia/Global/international/publications/toxics/Water%202011/dirty-laundry-report.pdf`, accessed February 14, 2018.

——— (2012): "Toxic Threads: Putting Pollution on Parade. How Textile Manufacturers are Hiding Their Toxic Trail," `http://www.greenpeace.org/eastasia/Global/international/publications/toxics/Water%202012/ToxicThreads02.pdf`, accessed February 14, 2018.

——— (2016): "The Detox Catwalk 2016. Who's On the Path to Toxic-Free Fashion?," `https://www.greenpeace.org/archive-international/en/campaigns/detox/fashion/detox-catwalk/`, accessed February 14, 2018.

Groden, C. (2015): "Hasbro, Mattel Toy Suppliers Slammed in Labor Report," `http://fortune.com/2015/11/22/hasbro-mattel-china-labor/`, accessed February 14, 2018.

Grossman, G. M., and E. Helpman (1994): "Protection for Sale," *American Economic Review*, 84(4), 833–850.

——— (2002): "Integration versus Outsourcing in Industry Equilibrium," *Quarterly Journal of Economics*, 117(1), 85–120.

Grossman, S. J., and O. D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4), 691–719.

Hainmueller, J., and M. J. Hiscox (2012): "The Socially Conscious Consumer? Field Experimental Tests of Consumer Support for Fair Labor Standards," *MIT Political Science Department Research Paper*, 2012-15.

Hallak, J. C. (2006): "Product Quality and the Direction of Trade," *Journal of International Economics*, 68(1), 238–265.

Harrison, A., and J. Scorse (2010): "Multinationals and Anti-sweatshop Activism," *American Economic Review*, 100(1), 247–273.

Hart, O., and J. Moore (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98(6), 1119–1158.

Hatte, S., and P. Koenig (2018): "The Geography of NGO Activism against Multinational Corporations," *World Bank Economic Review*, forthcoming.

Hiscox, M. J., and N. Smyth (2011): "Is There Consumer Demand for Fair Labor Standards? Evidence from a Field Experiment," *SSRN Working Paper*, 1820642.

Human Rights Watch (2004): "Turning a Blind Eye. Hazardous Child Labor in El Salvador's Sugarcane Cultivation," `https://www.hrw.org/sites/default/files/reports/elsalvador0604full_1.pdf`, accessed February 14, 2018.

Hummels, D., and P. J. Klenow (2005): "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95(2), 704–723.

Innes, R. (2006): "A Theory of Consumer Boykotts under Symmetric Information and Imperfect Competition," *Economic Journal*, 116, 355–381.

KantarMillwardBrown (2017): "Brandz Top 100 Most Valuable Global Brands 2017," `http://brandz.com/admin/uploads/files/BZ_Global_2017_Report.pdf`, accessed February 14, 2018.

King, B. G., and S. A. Soule (2007): "Social Movements as Extra-institutional Entrepreneurs: The Effect of Protests on Stock Price Returns," *Administrative Science Quarterly*, 52(3), 413–442.

Kitzmuller, M. (2012): "Is Better Information Always Good News? International Corporate Strategy and Regulation," *World Bank Policy Research Working Paper 6251*.

Kitzmuller, M., and J. Shimshack (2012): "Economic Perspectives on Corporate Social Responsibility," *Journal of Economic Literature*, 50(1), 51–84.

Kohler, W., and M. Smolka (2014): "Global Sourcing and Firm Selection," *Economics Letters*, 124(3), 411–415.

Krautheim, S., and T. Verdier (2016): "Offshoring with Endogenous NGO Activism," *Journal of International Economics*, 101, 22–41.

Locke, R. M., T. Kochan, M. Romis, and F. Qin (2007): "Beyond Corporate Codes of Conduct: Work Organization and Labour Standards at Nike's Suppliers," *Industrial Labour Review*, 146(1-2), 21–40.

Locke, R. M., F. Qin, and A. Brause (2007): "Does Monitoring Improve Labor Standards? Lessons from Nike," *Industrial Labor Relations Review*, 61(1), 3–31.

Loureiro, M., and J. Lotade (2005): "Do Fair Trade and Eco-Labels in Coffee Wake Up the Consumer Conscience?," *Ecological Economics*, 53(1), 129–138.

Lyon, T. P., and S. Salant (2013): "Linking Public and Private Politics: Activist Strategy for Industry Transformation," mimeo.

Melitz, M. J. (2003): "The Impact of Trade On Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Mosley, L. (2011): *Labor Rights and Multinational Production*, Cambridge Studies in Comparative Politics. Cambridge University Press.

Neslen, A. (2017): "Pepsico, Unilever and Nestlé Accused of Complicity in Illegal Rainforest Destruction," `https://www.theguardian.com/environment/2017/jul/21/pepsico-unilever-and-nestle-accused-of-complicity-in-illegal-rainforest-destruction`, accessed February 14, 2018.

Nunn, N. (2007): "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade," *Quarterly Journal of Economics*, 122(2), 569–600.

Nunn, N., and D. Trefler (2008): "The Boundaries of the Multinational Firm: An Empirical Analysis," in *The Organization of Firms in a Global Economy*, ed. by E. Helpman, D. Marin, and T. Verdier, pp. 55–83. Harvard University Press.

——— (2013): "Incomplete Contracts and the Boundaries of the Multinational Firm," *Journal of Economic Behavior and Organization*, 94, 330–344.

Nyborg, K. (2011): "I Don't Want to Hear about It: Rational Ignorance among Duty-Oriented Consumers," *Journal of Economic Behavior and Organization*, 79, 263–274.

O'Rourke, D. (2005): "Market Movements: Nongovernmental Strategies to Influence Global Production and Consumption," *Journal of Industrial Ecology*, 9(1), 115–128.

Ouellet, M., J. L. Short, and M. W. Toffel (2015): "Codes in Context: How States, Markets, and Civil Society Shape Adherence to Global Labor Standards," *Regulation and Governance*, 9, 205–223.

Pierce, J. R., and P. K. Schott (2012): "A Concordance Between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries," *Journal of Economic and Social Measurement*, 37(1-2), 61–96.

Rainforest Action Network (2017): "Leuser Watch - Destroying Elephant Habitat While Breaching the Indonesian Government Moratorium on Forest Clearance for Palm Oil," `https://d3n8a8pro7vhmx.cloudfront.net/rainforestactionnetwork/pages/17392/attachments/original/1486495650/RAN_Leuser_Watch_PT_Agra_Bumi_Niaga_20170207.pdf`, accessed February 14, 2018.

Rauch, J. E. (1999): "Networks versus Markets in International Trade," *Journal of International Economics*, 48, 7–35.

Ruwanpura, K. N., and N. Wrigley (2011): "The Costs of Compliance? View of Sri Lankan Apparel Manufacturers in Times of Global Economic Crisis," *Journal of Economic Geography*, 11, 1031–1049.

Schott, P. K. (2008): "The Relative Sophistication of Chinese Exports," *Economic Policy*, 53, 5–49.

Short, J. L. (2013): "Self-Regulation in the Regulatory Void: "Blue Moon" or "Bad Moon"?," *Annals of the American Academy of Political and Social Science*, 649(1), 22–34.

Smedley, T. (2014): "Sustainable Sugar: Coca-Cola and BP Signed Up But Will It Go Mainstream?," `https://www.theguardian.com/sustainable-business/2014/sep/15/sustainable-sugar-can-coca-cola-bp-shell-bonsucro`, accessed February 14, 2018.

Spar, D. L., and L. T. La Mure (2003): “The Power of Activism: Assessing the Impact of NGOs on Global Business,” *California Management Review*, 45(3), 78–101.

Spence, A. M. (1973): “Job Market Signaling,” *Quarterly Journal of Economics*, 87, 355–374.

——— (1974): *Market Signaling*. Harvard University Press.

Tomiura, E. (2007): “Foreign Outsourcing, Exporting, and FDI: A Productivity Comparison at the Firm Level,” *Journal of International Economics*, 72(1), 113–127.

Williamson, O. E. (1985): *The Economic Institutions of Capitalism*. Free Press.

Appendix

A1 Proof of Proposition 1

A1.1 Derivation of the supplier cutoff

For unethical production to be preferred, we need the total cost savings from unethical production ΔC to be larger than the expected ethical revenue premium $E\left[\Delta R_S\right]$.

$$\begin{gathered}\Delta C > E\left[\Delta R_S\right] \\ \left(c_m^e - c_m^u\right) m(\omega)_k^e > \left(1-\phi_k\right)\left(R(\omega)_k^e - E\left[R(\omega)_k^u\right]\right) \\ \left(c_m^e - c_m^u\right) m(\omega)_k^e > \left(1-\phi_k\right)\left(1-\gamma\right) R(\omega)_k^e \\ \left(c_m^e - c_m^u\right)\left(1-\beta\right) A\alpha^{\frac{1}{1-\alpha}} \frac{1-\phi_k}{c_m^e}\left[\left(\frac{c_h}{\phi_k}\right)^{\beta}\left(\frac{c_m^e}{1-\phi_k}\right)^{1-\beta}\right]^{-\frac{\alpha}{1-\alpha}} \\ > \left(1-\gamma\right)\left(1-\phi_k\right) A\alpha^{\frac{\alpha}{1-\alpha}}\left[\left(\frac{c_h}{\phi_k}\right)^{\beta}\left(\frac{c_m^e}{1-\phi_k}\right)^{1-\beta}\right]^{-\frac{\alpha}{1-\alpha}} \\ \frac{c_m^e - c_m^u}{c_m^e}\left(1-\beta\right)\alpha > 1-\gamma\end{gathered}$$

Solving for β using the fact that $c_m^u = \mu c_m^e$ gives that when

$$\beta < \beta_S = 1 - \frac{1-\gamma}{\left(1-\mu\right)\alpha}, \tag{24}$$

the supplier will prefer unethical production.

Comparative statics Differentiating w.r.t. $1-\mu$, $1-\gamma$, and $\frac{1}{\alpha}$ delivers

$$\frac{\partial \beta_S}{\partial\left(1-\mu\right)} = \frac{1-\gamma}{\alpha\left(1-\mu\right)^2} > 0. \tag{25}$$

$$\frac{\partial \beta_S}{\partial\left(1-\gamma\right)} = -\frac{1}{\alpha\left(1-\mu\right)} < 0. \tag{26}$$

$$\frac{\partial \beta_S}{\partial \frac{1}{\alpha}} = -\frac{1-\gamma}{\left(1-\mu\right)} < 0. \tag{27}$$

A2 Proof of Proposition 2

The cutoff β_l is the value of β that solves

$$\Theta^l(\beta_l) = \left[\left(\frac{\phi_V}{\phi_O}\right)^{\beta_l}\left(\frac{1-\phi_V}{1-\phi_O}\right)^{1-\beta_l}\right]^{\frac{\alpha}{1-\alpha}} \frac{\gamma - \alpha\,(1-\beta_l)\,\mu + \phi_V\alpha\,[\mu - \beta_l\,(1+\mu)]}{\gamma - \alpha\,(1-\beta_l)\,\mu + \phi_O\alpha\,[\mu - \beta_l\,(1+\mu)]} = 1 \tag{28}$$

with $\gamma, \mu \in (0,1)$ delivering the unethical cutoff β_u. In the corner case of $\gamma = \mu = 1$, the β that solves the equation is β_e.

A2.1 Existence

To show existence of the two cutoffs, we will derive conditions under which the corner cases $\Theta^l(\beta = 1) > 1$ and $\Theta^l(\beta = 0) < 1$ are true, implying that there exists some β_e for which $\Theta^e(\beta_e) = 1$ and some β_u for which $\Theta^u(\beta_u) = 1$.

Case 1: $\beta = 0$ $\quad\Theta^l(\beta)$ reduces to

$$\left(\frac{1-\phi_V}{1-\phi_O}\right)^{\frac{\alpha}{1-\alpha}}\left[\frac{\gamma - \alpha\mu\,(1-\phi_V)}{\gamma - \alpha\mu\,(1-\phi_O)}\right]. \tag{29'}$$

Case 2: $\beta = 1$ $\quad\Theta^l(\beta)$ becomes

$$\left(\frac{\phi_V}{\phi_O}\right)^{\frac{\alpha}{1-\alpha}} \frac{\gamma - \phi_V\alpha}{\gamma - \phi_O\alpha}. \tag{29''}$$

Here, again, $\gamma, \mu \in (0,1)$ deliver Θ^u and $\gamma = \mu = 1$ deliver Θ^e.

Numerator and denominator of each of the two cases differ only in the value of ϕ_k. Substituting x for $1-\phi_k$ in (29') and for ϕ_k in (29") and recalling that $\frac{1}{2} < \phi_k < 1$, the two cases only differ in the value of μ. Numerator and denominator of any of the two cases can be expressed in general form as

$$x^{\frac{\alpha}{1-\alpha}}\,(\gamma - \alpha\mu x)\,. \tag{29}$$

Because $\phi_V > \phi_O$ (and thus $1-\phi_V < 1-\phi_O$), conditions that ensure that equation (29) has a positive slope in x also ensure that $\Theta^l(\beta = 0) < 1$ and $\Theta^l(\beta = 1) > 1$.

$$\frac{\partial}{\partial x} x^{\frac{\alpha}{1-\alpha}}\,(\gamma - \alpha\mu x) = \frac{\alpha}{1-\alpha} x^{\frac{\alpha}{1-\alpha}}\left(\frac{\gamma}{x} - \mu\right)$$

Because $x \in (0,1)$ and $\alpha \in (0,1)$, the last factor determines the sign of the derivative. We must cover four cases, each of Cases 1 and 2 from above for ethical ($\gamma = \mu = 1$) and unethical production, i.e. with $\gamma, \mu \in (0,1)$.

$$\left.\begin{array}{r} \text{ethical production, } \beta = 0\text{: } \dfrac{1}{x} - 1 \\ \text{ethical production, } \beta = 1\text{: } \dfrac{1}{x} - 1 \\ \text{unethical production, } \beta = 0\text{: } \dfrac{\gamma}{x} - \mu \\ \text{unethical production, } \beta = 1\text{: } \dfrac{\gamma}{x} - 1 \end{array}\right\} \overset{!}{>} 0 \tag{30}$$

For ethical production, the condition always holds because $\frac{1}{x} > 1$ in both cases. To ensure existence of β_u, both conditions under unethical production must hold, i.e. we must have $\gamma > \mu\,(1-\phi_O)$ and $\gamma > \phi_V$. As $\gamma > \phi_V$ is the stricter condition, it is also a sufficient condition for existence.

Therefore, with ethical production, $\Theta^e(\beta = 1) > 1$ and $\Theta^e(\beta = 0) < 1$, therefore, β^e exists. With unethical production, if $\gamma > \phi_V$, then $\Theta^u(\beta = 1) > 1$ and $\Theta^u(\beta = 0) < 1$, therefore β^u exists. QED.

A2.2 Uniqueness

To establish uniqueness, we show under which conditions the derivative of $\Theta^l(\beta)$ with respect to β is larger than zero for all $\beta \in [0, 1]$. The proof follows the structure of Appendix 2 in Antràs (2003).

Recall that $\phi_V = \phi_O + \delta^\alpha (1 - \phi_O)$, where δ is the share of the intermediate the headquarter can continue to use in an integrated firm in case bargaining breaks down. Using this relationship, $\Theta^l(\beta)$ can be written as

$$\Theta^l(\beta) = \underbrace{\left[1 + \frac{\delta^\alpha}{\phi_O (1 - \delta^\alpha)}\right]^{\frac{\alpha\beta}{1-\alpha}} (1 - \delta^\alpha)^{\frac{\alpha}{1-\alpha}}}_{=F_1} \cdot \underbrace{\left[1 + \frac{\alpha\delta^\alpha (1 - \phi_O) [\mu - \beta (1 + \mu)]}{\gamma - \alpha (1 - \beta) \mu + \phi_O \alpha [\mu - \beta (1 + \mu)]}\right]}_{=F_2}.$$

As before, $\gamma, \mu \in (0, 1)$ deliver Θ^u and $\gamma = \mu = 1$ deliver Θ^e. The derivative of $\Theta^l(\beta)$ with respect to β is positive if

$$\Theta^{l\prime}(\beta) = \frac{\partial F_1}{\partial \beta} F_2 + \frac{\partial F_2}{\partial \beta} F_1 > 0,$$

with

$$\begin{aligned}
\frac{\partial F_1}{\partial \beta} &= (1 - \delta^\alpha)^{\frac{\alpha}{1-\alpha}} \ln\left(1 + \frac{\delta^\alpha}{\phi_O (1 - \delta^\alpha)}\right) \frac{\alpha}{1 - \alpha} \left[1 + \frac{\delta^\alpha}{\phi_O (1 - \delta^\alpha)}\right]^{\frac{\alpha\beta}{1-\alpha}} \\
\frac{\partial F_2}{\partial \beta} &= \frac{-\alpha\delta^\alpha (1 - \phi_O) (1 + \mu) [\gamma - \alpha (1 - \beta) \mu + \phi_O \alpha [\mu - \beta (1 + \mu)]]}{(\gamma - \alpha (1 - \beta) \mu + \phi_O \alpha [\mu - \beta (1 + \mu)])^2} \\
&\quad - \frac{\alpha\delta^\alpha (1 - \phi_O) [\mu - \beta (1 + \mu)] [\alpha\mu - \phi_O \alpha (1 + \mu)]}{(\gamma - \alpha (1 - \beta) \mu + \phi_O \alpha [\mu - \beta (1 + \mu)])^2}.
\end{aligned}$$

$\Theta^{l\prime}(\beta) > 0$ can be simplified to give

$$\ln\left(1 + \frac{\delta^\alpha}{\phi_O (1 - \delta^\alpha)}\right) \Omega (\beta, \mu, \gamma) > [\gamma (1 + \mu) - \alpha\mu] (1 - \alpha) (1 - \phi_O) \delta^\alpha$$

where

$$\Omega (\beta, \mu, \gamma) = \underbrace{[\gamma - \alpha\mu (1 - \phi_V) + \alpha\beta [\mu - (1 + \mu) \phi_V]]}_{\tau_V} \underbrace{[\gamma - \alpha\mu (1 - \phi_O) + \alpha\beta [\mu - (1 + \mu) \phi_O]]}_{\tau_O}.$$

The strategy is now to show that Ω strictly decreases in β and then to plug in the minimum value $\Omega (\beta = 1, \mu, \gamma)$ and show that the relationship still holds at this point. The two multiplicative terms τ_V and τ_O in Ω are symmetric except for the bargaining power parameter ϕ_k, so that

$$\frac{\partial \tau_k}{\partial \beta} = \alpha [\mu - (1 + \mu) \phi_k] < 0, \ k \ \in \ \{V, O\}.$$

To see this note that $\frac{\partial(\mu - (1+\mu)\phi_k)}{\partial \mu} = 1 - \phi_k > 0$. The term therefore reaches its maximum at $\mu = 1$, where it becomes $1 - 2\phi_k$, which is negative because $\phi_k > \frac{1}{2}$ by assumption. To determine the sign of $\frac{\partial \Omega}{\partial \beta}$, we need to determine the sign of τ_k, which can be rewritten as

$$\tau_k = \gamma - \alpha [\beta\phi_k + (1 - \beta) \mu (1 - \phi_k)].$$

The term in brackets can be shown to be smaller than ϕ_k because $\phi_k > \frac{1}{2}$. Therefore the assumption that $\gamma > \phi_V$ from the existence proof is sufficient to ensure a positive τ_k. Maintaining $\gamma > \phi_V$, it follows that under both ethical and unethical production, τ_k is positive. This implies that

$$\frac{\partial \Omega\left(\beta,\mu,\gamma\right)}{\partial \beta} = \frac{\partial \tau_V}{\partial \beta}\tau_O + \frac{\partial \tau_O}{\partial \beta}\tau_V < 0.$$

It follows that Ω attains its smallest value within the admissible range of β at $\beta = 1$. Plugging in $\beta = 1$ into Ω eliminates μ from the function and yields

$$\Omega\left(\beta = 1, \gamma\right) = \left(\gamma - \alpha\phi_V\right)\left(\gamma - \alpha\phi_O\right).$$

Note that the assumption $\gamma > \phi_V$ ensures that both factors are positive because $\phi_O < \phi_V$. Expressing ϕ_V in terms of ϕ_O and inserting this for Ω in $\Theta^l(\beta)$ and rearranging then yields

$$\vartheta(\delta) = \ln\left(1 + \frac{\delta^\alpha}{\left(1-\delta^\alpha\right)\phi_O}\right) - \frac{\left[\gamma\left(1+\mu\right) - \alpha\mu\right]\left(1-\alpha\right)\left(1-\phi_O\right)\delta^\alpha}{\underbrace{\left[\gamma - \alpha\left(\phi_O + \delta^\alpha\left(1-\phi_O\right)\right)\right]\left(\gamma - \alpha\phi_O\right)}_{=\Omega(\beta=1,\gamma)}} \overset{!}{>} 0$$

To show that $\vartheta(\delta) > 0$ for all $\delta \in (0,1)$, note that $\vartheta(\delta = 0) = 0$ so that $\vartheta(\delta) > 0$ if $\vartheta'(\delta) > 0$. The first derivative of ϑ with respect to δ can be expressed as

$$\begin{aligned}\frac{\partial \vartheta}{\partial \delta} =& \frac{\alpha\delta^{\alpha-1}}{\left(1-\delta^\alpha\right)\left[\delta^\alpha + \phi_O\left(1-\delta^\alpha\right)\right]} - \frac{\left[\gamma\left(1+\mu\right)-\alpha\mu\right]\left(1-\alpha\right)\left(1-\phi_O\right)}{\left(\gamma-\alpha\phi_O\right)\left[\gamma-\alpha\left(\phi_O+\delta^\alpha\left(1-\phi_O\right)\right)\right]^2} \\ &\cdot\left(\alpha\delta^{\alpha-1}\right)\left\{\left[\gamma-\alpha\left(\phi_O+\delta^\alpha\left(1-\phi_O\right)\right)\right]+\alpha\delta^\alpha\left(1-\phi_O\right)\right\} \overset{!}{>} 0.\end{aligned}$$

This can be simplified further to give

$$\left(\gamma - \alpha\phi_V\right)^2 \overset{!}{>} \left[\gamma\left(1+\mu\right) - \mu\alpha\right]\left(1-\alpha\right)\left(1-\phi_V\right)\phi_V \equiv M(\mu).$$

Now note that $\frac{\partial M}{\partial \mu} = \left(\gamma - \alpha\right)\left(1-\alpha\right)\left(1-\phi_V\right)\phi_V$. The sign of the derivative depends on the relationship between γ and α.

Case 1 Consider case 1 where $\gamma < \alpha$ and so $\frac{\partial M}{\partial \mu} < 0$. This implies that for $\mu \in (0,1)$, $M(\mu)$ attains a maximum in the corner case of $\mu = 0$. For the inequality above to hold it is therefore sufficient to prove that

$$\left(\gamma - \alpha\phi_V\right)^2 > \gamma\left(1-\alpha\right)\left(1-\phi_V\right)\phi_V. \tag{31}$$

Simplifying and solving for α equivalently gives

$$\alpha^2\phi_V^2 - \alpha\gamma\phi_V\left(1+\phi_V\right) + \gamma\left[\gamma - \left(1-\phi_V\right)\phi_V\right] > 0.$$

The discriminant term of this quadratic equation is given by

$$\left(1+\phi_V\right)^2\phi_V^2\gamma^2 - 4\phi_V^2\gamma\left[\gamma - \phi_V\left(1-\phi_V\right)\right].$$

Simplification shows that the discriminant term is negative if $\gamma > \frac{4\phi_V}{3+\phi_V}$ so that (31) has no roots and is thus always positive. Because $\frac{4\phi_V}{3+\phi_V} > \phi_V \ \forall \ \phi_V \in (0,1)$, the inequality (31) holds for all $\alpha \in (0,1)$ when $\gamma > \frac{4\phi_V}{3+\phi_V} > \phi_V$ and $\gamma < \alpha$.

We have previously imposed $\gamma > \phi_V$ to guarantee existence of β_u. Now consider values of γ between ϕ_V and $\frac{4\phi_V}{3+\phi_V}$. (31) has roots in this parameter range. For (31) to hold for all α for some $\gamma < \frac{4\phi_V}{3+\phi_V}$, we would need the smaller of the two roots of (31) to be larger than 1, which requires

$$\gamma\left(1+\phi_V\right)-2\phi_V>\sqrt{\left(1+\phi_V\right)^2\gamma^2-4\gamma\left[\gamma-\left(1-\phi_V\right)\phi_V\right]}. \tag{32}$$

The right-hand side is the discriminant term and is positive because we consider values of $\gamma<\frac{4\phi_V}{3+\phi_V}$. The left-hand side is only positive if $\gamma>\frac{2\phi_V}{1+\phi_V}$, which is larger than $\frac{4\phi_V}{3+\phi_V}$. This implies that in the range of values of γ we consider here, the left-hand side is always negative and so (32) never holds for these values. In the rest of the proof, we must therefore impose the stricter condition $\gamma>\frac{4\phi_V}{3+\phi_V}$.

Case 2 Consider case 2 where $\gamma>\alpha$ and so $\frac{\partial M}{\partial\mu}>0$. This implies that for $\mu\in(0,1)$, $M(\mu)$ attains a maximum at the corner case $\mu=1$. The relationship to be shown now is

$$\left(\gamma-\alpha\phi_V\right)^2-\left(2\gamma-\alpha\right)\left(1-\alpha\right)\left(1-\phi_V\right)\phi_V>0. \tag{33}$$

Note first that for the left-hand side to be increasing in γ, it has to hold that $\gamma>\phi_V\left[1-\phi_V\left(1-\alpha\right)\right]$. Because the term in brackets is smaller than 1, this is true for all $\gamma>\frac{4\phi_V}{3+\phi_V}\geq\phi_V$. It is therefore sufficient to show that (33) holds at the minimum level of γ. In this case we assume $\gamma>\alpha$ and impose $\gamma>\frac{4\phi_V}{3+\phi_V}>\phi_V$. Three sub-cases have to be covered.

Case 2a: $\gamma>\alpha>\frac{4\phi_V}{3+\phi_V}$ The minimum value γ can take here is α. Plugging in α for γ in (33) and simplifying gives that (33) holds when $\alpha>\phi_V$, which is true in this sub-case because $\frac{4\phi_V}{3+\phi_V}>\phi_V$.

Case 2b: $\gamma>\frac{4\phi_V}{3+\phi_V}>\alpha>\phi_V$ The minimum value γ can take here is $\frac{4\phi_V}{3+\phi_V}$. Case 2a has shown that if $\alpha>\phi_V$, (33) holds for $\gamma>\alpha$ which also holds in this case.

Case 2c: $\gamma>\frac{4\phi_V}{3+\phi_V}>\phi_V>\alpha$ Plugging in ϕ_V for γ in (33) results in the necessary condition of $\alpha<\phi_V$ for (33) to hold, which is true here. (33) therefore holds for $\gamma>\phi_V$ when $\phi_V>\alpha$. This includes $\frac{4\phi_V}{3+\phi_V}>\phi_V$. QED.

A2.3 Relative size of the two integration cutoffs

We prove that $\beta_u>\beta_e$ by showing that (1) $\frac{\partial\beta_u}{\partial\mu}<0$ for all $\mu\in(0,1]$ and $\gamma\in\left(\frac{4\phi_V}{3+\phi_V},1\right]$ and that (2) $\frac{\partial\beta_u}{\partial\gamma}<0$ for all $\mu\in(0,1]$ and $\gamma\in\left(\frac{4\phi_V}{3+\phi_V},1\right]$. This includes the corner case of $\mu=\gamma=1$, in which $\beta_u=\beta_e$. This implies that starting from the case $\beta_u=\beta_e$, any marginal decrease in either μ or γ increases β_u and continues to do so over the admissible range of the two parameters. We prove this using implicit differentiation of

$$\Theta^u(\beta_u)=F_1\cdot F_2=\left[1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right]^{\frac{\alpha\beta_u}{1-\alpha}}\left(1-\delta^\alpha\right)^{\frac{\alpha}{1-\alpha}}\cdot\left[1+\frac{\alpha\delta^\alpha\left(1-\phi_O\right)\left[\mu-\beta_u\left(1+\mu\right)\right]}{\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]}\right]=1$$

with respect to μ and γ.

Derivative of β_u with respect to μ First note that

$$\frac{\partial F_1}{\partial\mu}=\left(1-\delta^\alpha\right)^{\frac{\alpha}{1-\alpha}}\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\left[1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right]^{\frac{\alpha\beta_u}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{\partial\beta_u}{\partial\mu},$$

$$
\begin{aligned}
\frac{\partial F_2}{\partial \mu} &= \frac{\alpha\delta^\alpha\left(1-\phi_O\right)\left[1-\frac{\partial\beta_u}{\partial\mu}-\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)\right]\left[\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]\right]}{\left\{\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]\right\}^2} \\
&-\frac{\alpha\delta^\alpha\left(1-\phi_O\right)\left[\mu-\beta_u\left(1+\mu\right)\right]\left[\alpha\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)-\alpha+\phi_O\alpha\left[1-\frac{\partial\beta_u}{\partial\mu}-\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)\right]\right]}{\left\{\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]\right\}^2}
\end{aligned}
$$

and that $\frac{\partial 1}{\partial \mu}=0$. Combining the terms to write $\frac{\partial F_1}{\partial \mu}F_2+F_1\frac{\partial F_2}{\partial \mu}=0$ and simplification by multiplying through with the denominator term from $\frac{\partial F_2}{\partial \mu}$ gives that

$$
\begin{aligned}
&\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\Omega\left(\beta_u,\gamma,\mu\right)\frac{\partial\beta_u}{\partial\mu} \\
=&\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left\{\left[\alpha\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)-\alpha+\alpha\phi_O\left[1-\frac{\partial\beta_u}{\partial\mu}-\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)\right]\right]\left[\mu-\beta_u\left(1+\mu\right)\right]\right. \\
&\left.-\left[1-\frac{\partial\beta_u}{\partial\mu}-\left(\beta_u+\mu\frac{\partial\beta_u}{\partial\mu}\right)\right]\left[\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]\right]\right\},
\end{aligned}
$$

where $\Omega\left(\beta_u,\gamma,\mu\right)$ is defined as above. The term in braces can then be simplified and the expression becomes

$$
\begin{aligned}
&\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\Omega\left(\beta_u,\gamma,\mu\right)\frac{\partial\beta_u}{\partial\mu} \\
=&\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left\{\frac{\partial\beta_u}{\partial\mu}\left[\gamma\left(1+\mu\right)-\alpha\mu\right]-\gamma\left(1-\beta_u\right)-\alpha\beta_u^2\left(1-\mu\right)\right\},
\end{aligned}
$$

which can be rearranged to

$$
\begin{aligned}
&\frac{\partial\beta_u}{\partial\mu}\left[\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[\gamma\left(1+\mu\right)-\alpha\mu\right]-\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\Omega\left(\beta_u,\gamma,\mu\right)\right] \\
=&\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[\gamma\left(1-\beta_u\right)+\alpha\beta_u^2\left(1-\mu\right)\right].
\end{aligned}
$$

Notice that the term on the right-hand side is positive for the admissible ranges of the parameters. In particular, it is also positive for $\gamma,\mu\in(0,1]$. To get $\frac{\partial\beta_u}{\partial\mu}<0$, we need that the term in square brackets on the left-hand side is negative, or equivalently that

$$
\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[\gamma\left(1+\mu\right)-\alpha\mu\right]<\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\Omega\left(\beta_u,\gamma,\mu\right).
$$

Because we assume $\phi_k>\frac{1}{2}$, we know from the uniqueness proof in Section A2.2 that $\Omega\left(\beta_u,\gamma,\mu\right)$ has a minimum at $\beta_u=1$. Plugging in $\beta_u=1$ and rearranging shows that we need

$$
\ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)-\frac{\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[\gamma\left(1+\mu\right)-\alpha\mu\right]}{\Omega\left(\beta_u=1,\gamma,\mu\right)}\equiv\vartheta\left(\delta\right)>0
$$

to obtain $\frac{\partial\beta_u}{\partial\mu}<0$. In the uniqueness part in Section A2.2 above it has been shown that the condition above holds if $\gamma>\frac{4\phi_V}{3+\phi_V}$, and in particular this holds when $\mu=\gamma=1$. $\frac{\partial\beta_u}{\partial\mu}<0$ for $\mu\in(0,1]$ and $\gamma\in\left(\frac{4\phi_V}{3+\phi_V},1\right]$ implies that β_u is increasing in the *decide: unethical cost advantage* $1-\mu$ for any of these values of μ and γ.

Derivative of β_u with respect to γ First note that

$$\frac{\partial F_1}{\partial \gamma} = F_1 \ln\left(1+\frac{\delta^\alpha}{\phi_O\left(1-\delta^\alpha\right)}\right)\frac{\alpha}{1-\alpha}\frac{\partial \beta_u}{\partial \gamma}$$

and

$$\frac{\partial F_2}{\partial \gamma} = \frac{-\alpha\delta^\alpha\left(1-\phi_O\right)\left[\mu-\beta_u\left(1+\mu\right)\right]-\alpha\delta^\alpha\left(1-\phi_O\right)\left[\gamma\left(1+mu\right)-\alpha\mu\right]\frac{\partial\beta_u}{\partial\gamma}]}{\left(\gamma-\alpha\left(1-\beta_u\right)\mu+\phi_O\alpha\left[\mu-\beta_u\left(1+\mu\right)\right]\right)^2}.$$

Combining those two derivatives in the equation $F_1\frac{\partial F_2}{\partial\gamma}+F_2\frac{\partial F_1}{\partial\gamma}=0$ and solving for $\frac{\partial\beta_u}{\partial\gamma}$ gives

$$\frac{\partial\beta_u}{\partial\gamma} = \frac{\left(1-\alpha\right)\delta^\alpha\left(1-\phi_O\right)\left[\mu-\beta_u\left(1+\mu\right)\right]}{\ln\left(1+\frac{\delta^\alpha}{\phi_O(1-\delta^\alpha)}\right)\Omega\left(\beta_u,\mu,\gamma\right)-\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[\gamma\left(1+\mu\right)-\alpha\mu\right]}.$$

The sign of the derivative is ambiguous. The denominator is positive, including the case of $\mu=\gamma=1$, as can be seen from the uniqueness proof in Section A2.2. The numerator is only negative if $\beta_u>\frac{\mu}{1+\mu}$, where $\frac{\mu}{1+\mu}$ reaches its maximum of $\frac{1}{2}$ at $\mu=1$. Therefore, $\frac{\partial\beta_u}{\partial\gamma}<0$ iff $\beta_u>\frac{1}{2}$.

The strategy is now to show that $\beta_e>\frac{1}{2}$. This will then imply that when $\mu=\gamma=1$ and thus $\beta_u=\beta_e$, the numerator is negative and thus $\frac{\partial\beta_u}{\partial\gamma}<0$. This then proves that starting from $\beta_u=\beta_e$, any decrease in γ increases β_u and does so for the whole range of admissible parameter values, i.e. $\mu\in(0,1]$ and $\gamma\in\left(\frac{4\phi_V}{3+\phi_V},1\right]$. To see this, consider the parameter condition needed to produce $\beta_e=\frac{1}{2}$ as the ethical integration cutoff. $\Theta^e\left(\beta_e=\frac{1}{2}\right)=1$ after some algebra simplifies considerably to

$$\phi_O=D(\delta)\equiv\frac{1-\delta^\alpha}{2-\delta^\alpha}.$$

As $\delta,\alpha\in(0,1)$, $D(\delta)$ reaches its maximum of $\frac{1}{2}$ as $\delta\to0$. This means that to have $\beta_e=\frac{1}{2}$, we need $\phi_O=D(\delta)$ with $D(\delta)<\frac{1}{2}$. This is ruled out by the initial assumption that $\phi_O>\frac{1}{2}$, which we carry over from Antràs (2003). We have now merely shown that $\beta_e=\frac{1}{2}$ is impossible under the imposed parameter restrictions. The proof is only complete if we show that any $\beta_e<\frac{1}{2}$ requires a value of ϕ_O whose maximum also lies below $\frac{1}{2}$. Therefore, we show that $\frac{\partial\beta_e}{\partial\phi_O}>0$, implying that a decrease in β_e requires a reduction in ϕ_O, c.p. Implicit differentiation yields that

$$\frac{\partial\beta_e}{\partial\phi_O} = \frac{\frac{\Omega(\beta_e)\beta_e\delta^\alpha}{\phi_O[\phi_O(1-\delta^\alpha)+\delta^\alpha]}+\delta^\alpha\left(1-\alpha\right)\left[1-2\beta_e\right]\left[1-\alpha\beta_e\right]}{\ln\left(1+\frac{\delta^\alpha}{\phi_O(1-\delta^\alpha)}\right)\Omega\left(\beta_e\right)-\delta^\alpha\left(1-\phi_O\right)\left(1-\alpha\right)\left[2-\alpha\right]}.$$

The sign of the derivative is again determined by the sign of the numerator. The denominator is positive as has been shown in the uniqueness part of the proof in Section A2.2. If $\beta_e\leq\frac{1}{2}$, the numerator is positive. Therefore, for $\beta_e\leq\frac{1}{2}$, a marginal decrease in β_e would require a decrease in ϕ_O. So in order to have a $\beta_e<\frac{1}{2}$ we require $\phi_O<D(\delta)$, which is ruled out by the initial assumption of $\phi_O>\frac{1}{2}$. QED.

A3 Proof of Proposition 3

For the proofs of Cases 2 and 3 note that the existence part of the proof of Proposition 2 specifies conditions for which $\Theta^u(\beta_u)$ and $\Theta^e(\beta_e)$ are *smaller* than 1 and *larger* than 0, respectively. Therefore, as long as these conditions hold, $\beta_e\in(0,1)$ and $\beta_u\in(0,1)$. Showing that $\beta_S\leq0$ and $\beta_S=1$ are possible within the admissible range of the parameters determining the cutoff proves the existence of Cases 2 and 3. With these preliminaries, it is unnecessary to consider partial derivatives of β_e and β_u with respect to μ and γ, because by Proposition 2, $\beta_e,\beta_u\in(0,1)$.

A3.1 Case 3

For this case we show that as the unethical cost advantage goes to zero ($\mu \to 1$), $\beta_S \to -\infty$ so that unethical production is never chosen.

$$\lim_{\mu \to 1} \beta_S = \lim_{\mu \to 1} \left[1 - \frac{1-\gamma}{\alpha (1-\mu)} \right] = -\infty. \tag{34}$$

A3.2 Case 2

For this case we show that $\beta_S \to 1$ as the threat of a consumer boycott goes to zero ($\gamma \to 1$) so that ethical production is never chosen.

$$\lim_{\gamma \to 1} \beta_S = \lim_{\gamma \to 1} \left[1 - \frac{1-\gamma}{\alpha (1-\mu)} \right] = 1 - \frac{0}{\alpha (1-\mu)} = 1. \tag{35}$$

A3.3 Case 1

Consider some $\beta_S \in (-\infty, 1)$. Case 1 trivially exists if $\beta_e < \beta_S < \beta_u$. Case 1 also exists starting from any value of $\beta_S < \beta_e$ or $\beta_S > \beta_u$. If $\beta_S < \beta_e$, increasing $\gamma \to 1$ will necessarily move $\beta_S \to 1$, while $\beta_e, \beta_u \in (0,1)$. For some values of γ given μ and α, it must be the case that $\beta_e < \beta_S < \beta_u$. If $\beta_S > \beta_u$, increasing $\mu \to 1$ will necessarily move $\beta_S \to -\infty$, while $\beta_e, \beta_u \in (0,1)$. For some values of μ given γ and α, it must be the case that $\beta_e < \beta_S < \beta_u$. QED.

A4 Proof of Proposition 4

It was shown in the proof of Proposition 2 in Section A2.3 that $\frac{\partial \beta_u}{\partial \mu} < 0$ for $\mu \in (0,1]$ and $\gamma \in \left(\frac{4\phi_V}{3+\phi_V}, 1 \right]$. It follows directly that $\frac{\partial \beta_u}{\partial (1-\mu)} > 0$ for these parameter values. It has been shown in the proof of Proposition 1 in Section A1 that $\frac{\partial \beta_S}{\partial (1-\mu)} > 0$. Moreover, it can be seen from equation (19) that β_e does not depend on μ or γ. Therefore, $\frac{\partial \beta_e}{\partial \mu} = 0$. QED.

A5 Proof of Proposition 5

The proof follows closely the proof of Proposition 1 in terms of structure. When the headquarter can also set the technology of the match in addition to the organizational form, the key difference is that the headquarter takes the overall surplus into account when deciding between ethical and unethical production. The headquarter again compares the total cost savings from unethical production ΔC to the expected ethical revenue premium, which we now label $E[\Delta R]$, which is given by the sum of the suppliers and the headquarters revenue premium. Therefore, the term $(1-\phi_k)$ on the right-hand side, which denoted the revenue share allocated to the supplier in the proof of Proposition 1 is now replaced by unity.

$$
\begin{aligned}
\Delta C &> E\left[\Delta R\right] \\
\left(c_m^e - c_m^u\right) m(\omega)_k^e &> \left(R(\omega)_k^e - E\left[R(\omega)_k^u\right]\right) \\
\left(c_m^e - c_m^u\right) m(\omega)_k^e &> \left(1-\gamma\right) R(\omega)_k^e \\
\left(c_m^e - c_m^u\right)\left(1-\beta\right) A\alpha^{\frac{1}{1-\alpha}} \frac{1-\phi_k}{c_m^e} \left[\left(\frac{c_h}{\phi_k}\right)^{\beta}\left(\frac{c_m^e}{1-\phi_k}\right)^{1-\beta}\right]^{-\frac{\alpha}{1-\alpha}} & \\
> \left(1-\gamma\right) A\alpha^{\frac{\alpha}{1-\alpha}} \left[\left(\frac{c_h}{\phi_k}\right)^{\beta}\left(\frac{c_m^e}{1-\phi_k}\right)^{1-\beta}\right]^{-\frac{\alpha}{1-\alpha}} & \\
\frac{c_m^e - c_m^u}{c_m^e}\left(1-\phi_k\right)\left(1-\beta\right)\alpha > 1-\gamma &
\end{aligned}
$$

Solving for β using the fact that $c_m^u = \mu c_m^e$ gives that when

$$
\beta < \beta_{H,k} = 1 - \frac{1-\gamma}{\left(1-\mu\right)\alpha\left(1-\phi_k\right)} < \beta_S, \tag{36}
$$

the headquarter will prefer unethical production. Note that this cutoff now depends on the organizational form of the firm. $\beta_{H,O} > \beta_{H,V}$ because $\phi_V > \phi_O$. Because $1-\phi_k < 1$, both cutoffs are smaller than β_S from the baseline model.

Existence of the described pattern From Section A3 we know that by letting $\gamma \to 1$, $\beta_S \to 1$ and by letting $\mu \to 1$, $\beta_S \to -\infty$. Because the new cutoffs $\beta_{H,k}$ with $k \in \{V, O\}$ differ from β_S only by a positive factor in the denominator, the results from Proposition 3 can be directly applied to the cutoffs derived above. Therefore, there is a non-empty set of admissible values of γ, α and μ that ensures that for any $\phi_k > \frac{1}{2}$, there exists a range of $\beta \in (\beta_{H,O}, \beta_S) \wedge \beta \in (\beta_e, \beta_u)$. QED.