Essays on Microeconomic Theory

Jasmin Droege
Nuffield College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics

Trinity 2022

Wordcount: 90,099

Für mein Patenkind, dessen sehnlichst erwartete Ankunft am 03.06.2022 meine Welt auf den Kopf gestellt hat.

To my godchild, whose eagerly awaited arrival on 03.06.2022 has turned my world upside down.

Acknowledgements

I would like to take this opportunity to thank those people without whom this work would not have been possible. I owe a great deal of thanks to my initial Supervisor Dr Willemien Kets for her guidance in developing the three lines of research in this thesis, and her invaluable support to get my work out there and emerge stronger from every rejection I received over the years. A special word of thanks goes to Dr Séverine Toussaert for taking me under her wing upon my return to the University of Oxford in Trinity 2022. Without her, I would not have reached the finish line. Finally, I am deeply thankful to Dr Meg Meyer for her guidance in her role as College Supervisor.

I have had the privilege of getting two of my essays out there during my DPhil journey at the University of Oxford. This would not have been possible without the dedication of the Editor in Chief Professor Douglas S. Noonan and two anonymous referees at the *Journal of Cultural Economics*, as well as the General Editor Professor Lawrence J. White and an anonymous referee at the *Review of Industrial Organization.*

I would like to thank Nuffield College, especially the Patrons of the Terence Gorman Scholarship, and the Economic and Social Research Council [grant number ES/P000649/1] for providing the financial support necessary for me to undertake the DPhil. In conjunction with this, I am particularly thankful to Nuffield's Senior Tutor Dr Eleni Kechagia-Ovseiko, for assisting me in making the DPhil fit around my needs and aspirations and for her continual support on behalf of the College during my suspension of status.

I am indebted to Dr Matthew J. Robertson for being my significant other, co-author and provider of countless proofreading marathons. Last but not least, I would like to thank my Twin Sister, aka my Doppelgänger, and my Parents for all their support and faith in me. Without you, I would not have reached Oxford in the first place.

Jasmin Droege
Oxford, July 2022

Contents

List of Figures

Abstract

In three distinct, yet interrelated essays I examine real-world phenomena at the intersection of microeconomic theory and behavioural economics. To do so, I modify longstanding concepts in information economics, including signalling and games of incomplete information, to include *non-standard preferences*, *non-standard beliefs*, and *non-standard decision-making.* As a result, I obtain novel theoretical predictions that provide insights for policy.

In the **first** essay, I develop a game-theoretic framework to study the repercussions of an evaluator's bias against a specific group of applicants. The evaluator wants to hire a highly able applicant who delivers a high-quality performance; at the same time, he prefers hiring a male over a female. The evaluator decides upfront between holding an informed or a blind audition. In the latter, the evaluator learns neither the applicant's ability nor the gender. I show that, above a threshold bias, the evaluator prefers a blind audition to provide high effort incentives exclusively for high-ability applicants. Consequently, committing to no information can be beneficial for the evaluator. I also show that a highly biased evaluator's preferences align with those of a highly able female. I extend the framework to performance uncertainty and gender-blind CVs, and compare blind auditions to affirmative action. The framework is relevant for auditory-based applications: my results can explain why blind auditions have increased the probability of a female orchestra musician being hired via taste-based discrimination and challenge explanations grounded in statistical discrimination.

Although online marketplaces for handmade products persist, little theoretical research has been undertaken to explain why firms choose a handmade strategy. In the **second** essay, I develop a model that can explain the persistence through a handmade effect on the consumer side. I show that when consumers are willing to pay a sufficiently high handmade premium, the monopolistic firm chooses production by hand over superior machine production. When the firm is part of a duopoly, the existence of consumers who care about the conditions under which a product is manufactured can explain the firms' specialisation and,

thus, the observed co-existence of handmade and machine-made products in the economy. Such specialisation is efficient, and can be robust to collusion. The presence of shoppers who are uncertain about the appropriate behaviour may enable the monopolist to use a handmade strategy to signal a social norm of conscious consumption.

In the **third** essay, we study heuristic decision-making in social settings without communication. Players trade off conforming to the average action and matching their private type, which is correlated with the state of the world. We contrast individual and aggregate behaviour under two heuristics as well as rationality: *credulous* players posit others truthfully reveal their types whilst *sceptical* players reflect on predecessors' actions, discounting any action in the history of play that could be a falsification; in contrast, *rational* players take into account the extent to which others wish to conform and may, thus, not choose to truthfully reveal. We apply our results to organisational decision-making and social engineering, and show that, by making a credulous player sceptical, a policy maker can increase truth telling. Making a player rational, in contrast, does not guarantee that more players truthfully reveal. At the aggregate level, rationality is not only necessary to incentivise players to adopt the action matching the state, but also preferable given a utilitarian objective. Making the state common knowledge can be a viable alternative whenever a policy maker prefers truth telling if and only if a player's type matches the state.

Chapter 1

Introduction

A fundamental tenet of microeconomic theory is that information has economic value: individuals increase their expected payoffs when making more informed decisions. While it tends to hold that better information is more costly to acquire, many aspects of the "algebra of information" differ from that of ordinary goods – complicating the economic analysis through, inter alia, increasing returns to information or the cheap reproduction of pieces of information following their production (Arrow, 2017, p.126). Amid those complications, the development of an *economics of information* (Stigler, 1961) over the latter half of the twentieth century has boosted the relevance of the profession to society in a number of ways.

First, situations in which one party has more information than the other are ubiquitous in the economy, and the profession's study of such information asymmetries has provided insights as to when we would expect this to lead to a change in behaviour and what inefficiencies might arise through adverse selection or moral hazard (as in my **first** essay). While adverse selection characterises a situation when the better-informed party selectively participates in an economic exchange at the expense of the less informed party, moral hazard can arise when a more informed party (the agent) acts on behalf of another less informed party (the principal). If their incentives do not align, the more informed party tends to act in a too risky manner from the viewpoint of the less informed party. Akerlof (1970) first examined the role of adverse selection in the market for used cars, illustrating how a market can fail when buyers and sellers have different information. Holmström (1979) considered the role of moral hazard in contracting and how the parties settle contractual arrangements amid fears of free riding of the more informed party.

The profession has not only devised a formalisation of information asymmetries and a framework to analyse their implications for economic exchange, bargaining and contracting, it has also equipped policy makers with ingenious tools for solving inefficiencies, such as market failure, arising from information asymmetries. Steps to eliminate or mitigate the impact of informational asymmetries include signalling (used in my **first** and **second** essay) and screening, consumer protection laws, and warranties, as well as the allocation of control, property, or decision rights between parties. Spence (1973) first showed that under certain conditions, the more informed party can signal their private information to the less informed party to improve the outcome for both sides. Stiglitz (1975) showed that a less informed party can screen the more informed party; that is, capture the information of the more informed party by providing choices from a menu of contracts.

A second way in which information economics has boosted the relevance of the profession to society is the insight that economic situations can often be approximated as games of incomplete information. This is because a variety of features of the economic environment are often not commonly known: the parties engaging in an economic exchange may be uncertain about the others' payoffs or their identity, what the other parties know, what actions they can take or how outcomes depend on the action taken. Amid uncertainty about the structure of the game, the parties can, however, oftentimes form some beliefs about these features; for example, based on their own longstanding experience. Harsanyi (1967, 1968a,b) recognised this as an opportunity to apply aspects of Bayesian probability. He considered how parties make decisions in these situations by modelling nature as a player that chooses the game's parameters upfront according to some commonly known probability distribution. Instead of maximising payoffs, this new player uses a fixed mixed strategy. While nature remains a theoretical construct in these games to incorporate, for example, the players' private information or types (used in my **first** and **second** essay) or the state of the world (used in my **third** essay), it converts an incomplete information game to one of complete, but imperfect, information. Among the seminal papers in information economics, Harsanyi's works seize a unique place due to the generality of his approach, which can be applied to the economic modelling of various informational problems (Myerson, 2004).

1.1 From Homo Economicus to Behavioural Agents

While the profession developed information economics for a rigorous study of information asymmetries, tools for solving inefficiencies arising from these asymmetries, as well as games of incomplete information, the field failed to recognise that human reasoning need not conform to statistical principles - including the effects of sample size, correlation or base rate. An individual's statistical intuition can persistently differ from their statistical knowledge (Kahneman, 2002). What is more, they may simplify the complex task of assessing probabilities and predicting values through heuristic rules, such as representativeness, availability, and anchoring. Specifically, the availability heuristic (used in my **third** essay) is a rule-of-thumb that relies on the memory of specific instances that are easily recollected – usually something that happened recently or looms large in one's mind. While heuristics are economical and usually effective because they allow us to decide quickly, they come at the risk of inducing bias; that is, settling on a suboptimal or inefficient outcome due to systematic and predictable errors in an individual's estimation (Tversky and Kahneman, 1974). In music, for example, adjudicators have been shown to be "blinded by their sight": their performance evaluation relies primarily on visual cues while it is the sound that is agreed to be central to the evaluation (Tsay, 2013, 2014). Such a vision heuristic (used in my **first** essay), thus, risks overlooking a highly qualified candidate because the musician's perceived ability is affected by their appearance (Griffiths, 2008).

It took the economics profession three decades to formally recognise the contributions of Tversky and Kahneman's (1974) "heuristics and biases approach" to an applied economic modelling of decision-making under uncertainty. Only in October 2002 did psychologist Daniel Kahneman join the ranks of renowned economists to win the Nobel memorial Prize in Economics for his works that challenge the assumption of human rationality in economic theory. By applying psychological insights to judgment and decision-making under uncertainty, he transformed our understanding of how people actually make economic decisions. His research served as a reality check to the orthodox neoclassical economic theory and revealed the need for the formulation of a new branch of economics. This new branch, called *behavioural economics*, builds on the existing framework but also models the human limits to rationality and the need to use "fast and frugal" heuristics in light of limited information-processing abilities (Gigerenzer et al., 1999).

Richard Thaler stepped into Tversky and Kahneman's seminal work, advancing the study of simple rules-of-thumb or heuristics to make decisions. He took the risk to pursue a new way of doing economics which entertains the possibility that not everyone in the economy is necessarily rational - that people may only be boundedly rational (Simon, 1955). This included advocating for an economic theory of the consumer that does not only specify how a consumer *should* choose but also describes how they *do* choose. Such a descriptive theory allows the consumer to underweight opportunity costs, be affected by sunk costs, search until the expected amount saved as a proportion of the total price of the purchase equals some critical value,[1] voluntarily restrict their choices to avoid regret, or precommit to solve self-control problems (Thaler, 1980). Through his regular column on 'Anomalies' in the *Journal of Economic Perspectives*, Thaler helped popularise behavioural economics among young readers and established economists alike.[2]

[1]Rather than searching for additional prices until the expected saving from the purchase equals the cost of visiting one more store (Stigler, 1966).

[2]According to Thaler (2015, p.258), the essential elements of behavioural economics are "the three bounds": bounded rationality, bounded willpower, and bounded self-interest.

Of particular interest is the insight that "[p]eople like to do what most people think is right to do; people like to do what most people actually do" (Thaler and Sunstein, 2008, p.182). It has profound implications for our understanding of both consumer decision-making at the microeconomic level, as well as fashions and fads at the macroeconomic level. First, herd mentality can determine a consumer's buying habits. For example, technologies are oftentimes experience products. Therefore, an individual may find it difficult to determine a new gadget's benefits. Faced with uncertainty, they may, thus, decide to use a heuristic or shortcut of discounting their own beliefs and instead imitate their friends in the adoption decision of Amazon's Kindle, Apple's iPhone, as well as Web 2.0 technologies (Sun, 2013). Second, herd mentality can explain why the sales of a product skyrocket even when not many people (intend to) actually use it. Examples include the Hula Hoop toy fad in the 1950s with an estimated 20 million lightweight plastic circles sold in the first six months of production alone (Internicola, 2009), or "Tupperware parties" in the 1960s and 1970s where housewives gathered together to purchase an increasingly unnecessary mountain of Tupperware - failing to resist the 'burp & seal' feature of these plastic containers (Nostalgia Central, n.d.). What this discussion suggests is that organisations tap into the insights from behavioural economics and understand that they need to create perceptions about their products (such as the party-plan selling strategy of Tupperware) to set a trend (used in my **second** essay).

1.2 My Contribution

In my thesis, I build on longstanding concepts in microeconomic theory and information economics, as well as the recent branch of behavioural economics to examine real-world phenomena through a different lens: in my first two essays, I combine, inter alia, the concept of signalling (Spence, 1973) and Lancaster's (1966) product characteristics approach with *non-standard preferences*; my third essay combines the theoretical framework of preference falsification (Kuran, 1987a,b, 1995) with *non-standard beliefs* and *non-standard decision-making.*[3] By introducing deviations from the standard models in various steps of the decision-making process, I obtain novel predictions that provide insights for policy.

My **first** essay is inspired by the observation that many U.S. symphony orchestras moved to blind auditions in the 1970s and 1980s (Goldin and Rouse, 2000). The use of curtains has been marketed as a fairer form of audition and equal opportunity policy, protecting women and minorities against discrimination (Karl Schiebler in The Economist, 1996). The

[3] DellaVigna (2009) argues that behavioural economics considers three classes of deviations from the standard theory: non-standard preferences, non-standard beliefs, and non-standard decision-making.

practice is often credited to have its origins in 1970 with the request of the late African American bassist Art Davis for candidates to play behind a curtain when auditioning for the New York Philharmonic. The increasing focus on equal opportunity in the performing arts has spawned more unconventional entertainment formats. The singing competition franchise "The Voice", adapted in 129 regions across the world, uses blind auditions in the first round. When the judges listen, they face the audience rather than the contestant. Only if a judge commits to taking on the contestant can she press a button which turns her chair around. According to creator John de Mol, this allows the judges to focus on quality alone and identify genuine singing talent (Universal Music Group, 2011).

In this essay, I address the two polarised views on the use of blind auditions to hire the most talented candidate. Proponents like Gladwell (2005) argue that blind auditions have helped evaluators avoid misleading first impressions. He bases his argument on the prejudices that women have encountered in the past when auditioning for orchestra positions. Two examples of such are not possessing the physical strength for instruments traditionally considered "male", or an inferior attitude and resilience compared to males. Opponents fear that, while avoiding bias, the evaluator also loses valuable information that can screen applicants by ability. As a result, top talent may be lost in an increasingly large applicant pool and standards of fairness may actually deteriorate (Holland, 1981).

I develop a game-theoretic framework to study the repercussions of an evaluator's bias against a specific group of applicants. The evaluator decides upfront between holding an informed or a blind audition. Only in the former can the evaluator learn the applicant's ability and their gender. For an informed audition, I draw upon the literature on cognitive biases, arguing that the evaluator is likely subject to different types of biases, summarised in a reduced-form parameter as a *non-standard preference*. Potential drivers are intergroup bias (Hewstone et al., 2002; Anwar, 2012; Bagues and Perez-Villadoniga, 2012), gender associations with instruments (Abeles, 2009; Stronsick et al., 2018; Sergeant and Himonides, 2019), and extraneous factors like attractiveness or accent (Purkiss et al., 2006). At the same time, I deploy longstanding concepts from information economics. An informed audition is characterised by verifiable disclosure (Milgrom, 1981; Grossman, 1981) as the applicant cannot misrepresent their gender or ability on a CV. A blind audition is akin to the signalling model of Spence (1973) where the applicant's effort decision may be informative about the ability dimension of their type while gender cannot be signalled.

I show that, above a threshold bias, the evaluator prefers a blind audition to provide high effort incentives exclusively for high-ability applicants. Consequently, committing to no information can be beneficial for the evaluator. Moreover, above this threshold bias, there is no trade-off between implementing blind auditions as a policy to counteract first impression biases and efficiency due to adverse effects on targeted effort provision. Perhaps surprisingly, a highly biased evaluator's preferences align with those of a highly able female. I extend the framework to a situation in which an applicant's performance quality is stochastically determined by effort and one additional uncertainty parameter, as well as as the option to request a gender-blind CV that contains valuable information about ability prior to a blindfolded performance, and compare blind auditions to affirmative action.

My framework is relevant for auditory-based applications: my results can explain why blind auditions have increased the probability of a female orchestra musician being hired via taste-based discrimination. However, going forward, blind hiring could be adapted in industries other than the performing arts; for example, to address the lower success rates of female applicants for funding schemes such as ERC grants (Vernos, 2013). The technology company GapJumpers, for example, offers blind hiring to firms in preliminary rounds. The Silicon Valley start up has developed software that withholds information about ethnicity, gender, age or educational background. An online test tailored to the skills in the job specification allows candidates to essentially blind-audition for a job, and employers to signal that "[i]f you can show me your skills in this role, I am willing to interview you, regardless of where you come from, what you look like or who you are" (CEO Kedar Iyer in Cain Miller, 2016).

Policy makers can learn from my framework that – contrary to conventional wisdom – equal opportunity and efficiency need not be conflicting policy objectives. Under certain conditions, blind auditions as an equal opportunity policy outperform more drastic policies, such as quotas and subsidies, aimed at mitigating first impression biases. The framework provides policy makers with the insight that blind auditions may need to be coupled with ability-targeting policies when the environment is characterised by randomness in order to stimulate the acquisition of skills and prevent market failure.

My **second** essay is inspired by the success of Etsy's "handmade" business model in the twenty-first century. Founded only in 2005 as a marketplace for handmade products, Etsy registered its one-millionth sale just two years later (Walker, 2007). Everything listed in the marketplace must be "handmade or unique and assembled with production partners, vintage or craft supplies" (Etsy, Inc., 2019a, p.11). Hence, Etsy can essentially be described as an infinitely large virtual craft fair, open 24/7 to anyone anywhere in the world, which allows consumers to participate in the DIY movement even if they do not have the enthusiasm to get crafty themselves (Walker, 2007). The success of Etsy underlines that crafts in the twenty-first century are being rediscovered as a desirable enterprise (Jakob, 2012); going so far that the UK government since 2010 has been deploying creative industries policies to revitalise local manufacturing through the promotion of craftsmanship and hand skills (Jakob and Thomas, 2017). However, amid Etsy's success, little theoretical research has been undertaken to explain why firms choose a handmade strategy in an era of technological advancement, which would allow for automated production processes in a wide range of categories - including jewellery, clothing, toys and furniture - potentially resulting in improved quality and fewer defective products.

In this essay, I develop a model that can explain the persistence of this traditional sector of the economy through a handmade effect on the consumer side. Shoppers in my model come in two types: quality shoppers, and conscious shoppers. I assume that all shoppers value product quality equally but they differ in the weight given to the production process. In assuming that conscious shoppers have *non-standard preferences*, I build upon the growing economic literature on ethical consumption where a group of consumers have preferences over how a good is produced and distributed rather than its physical properties alone (e.g., Stiefenhofer, 2019), and the empirical microeconomics literature that estimates consumers' willingness to pay a premium for ethical products (e.g., Galarraga and Markandya, 2004; Loureiro and Lotade, 2005; Arnot et al., 2006). At the same time, my assumption draws upon Lancaster's (1966) well-known microeconomic demand theory as shoppers do not choose between different goods but between the characteristics they provide. As in my first essay, I deploy longstanding concepts from information economics when I extend the baseline model to a signalling model in the spirit of Sliwka (2007): conformist shoppers as a third shopper type are influenced by social norms because they are uncertain about the appropriate consumption behaviour.

In the baseline model, I show that when consumers are willing to pay a sufficiently high handmade premium, the monopolist chooses production by hand over superior machine production, irrespective of whether hand production is also characterised by constant or instead characterised by increasing per-unit cost. If not limited to one product for a credible handmade effect, the monopolist adds a handmade product line in this case. When the firm is part of a duopoly, the existence of consumers who care about the conditions under which a product is manufactured can explain the firms' specialisation and, thus, the observed coexistence of handmade and machine-made products in the economy. Such specialisation is efficient, and can be robust to collusion. Finally, I show in the extension that the presence of shoppers who are uncertain about the appropriate behaviour may enable a monopolist to use a handmade strategy to signal a social norm of conscious consumption.

My model can shape our intuitions on ethical production and the development of policy that fosters conscientious consumption behaviour. First, firms rationally weigh the additional costs against the expected benefits when deciding whether to sell, say, a Fair-Trade product. The Fair-Trade premium that the firms can charge is the consumers' maximum willingness to pay for the subjective services of the Fair-Trade label. Moreover, premium-priced Fair-Trade products need not deliver any extra physical quality to co-exist alongside conventional products in the economy, and their label is crucial for market segmentation and third-degree price discrimination. Finally, building positive attitudes and removing scepticism towards Fair Trade can positively impact conscientious buying behaviour. The existence of impressionable consumers that are influenced by social consumption norms can explain the rising demand for and the mainstreaming of Fair Trade in Europe in the late 1990s (Renard, 2003) and social movements such as "buyhandmade.org" (Walker, 2007) through a crowding-in effect.

My **third** essay is inspired by a long-known insight from psychology: as actors, individuals tend to attribute their own behaviour to situational causes, but as observers, explain the identical behaviour of others through their inherent qualities or traits (Jones and Nisbett, 1971; Nisbett et al., 1973).[4] This might be familiar to any teacher of intermediate microeconomics. An intuitive approach to introducing the concept of price elasticity of demand can lead to misunderstandings in the case of linear demand (Daskin, 1992); however, a calculus-based approach requires some degree of mathematical sophistication (Wei, 2013). Hence, under either approach, some students in the classroom are likely to require further explanations to understand the material. A mindful professor may, therefore, ask the class if there are any clarifying questions. Rather than being inundated with questions, however, the typical response from students is silence. Students tend to interpret their peers' silence heuristically at face value as a signal that their peers have understood the concept. The heuristic causes students to ignore the possibility that they are not the only one requiring a different approach to understand the concept, so that ultimately the heuristic paralyses the whole class from asking their burning questions.

In this essay, we study the implications of such heuristic decision-making in social settings without communication. We set out a model in which players trade off conforming to the average action and matching their private type, which is correlated with the state of the world. After being sorted into a random order, each player chooses a publicly observable action given her type and the history of actions. We contrast individual and aggregate behaviour under two heuristics as well as rationality: *credulous* players posit others truthfully reveal their types whilst *sceptical* players reflect on predecessors' actions to edit the history of play; in contrast, *rational* players take into account the extent to which others wish to conform. Borrowing Thaler's (2015) terminology, we essentially compare what an "Econ"

[4]This actor-observer difference is related to, but distinct from, a self-serving bias whereby the actor would attribute cause for success to themselves but cause for failure to situational factors. In light of a self-serving bias, we would expect an actor-observer difference primarily after failure (Gioia and Sims Jr., 1985).

would do with what a "Human" would do, assuming they (i) have *non-standard beliefs* in the sense of overestimating the extent to which some or all players are concerned with being truthful to their private type, and (ii) make *non-standard decisions* in the sense of using a heuristic that discounts any action in the history of play that could be a falsification.

Our specification draws upon experimental literature in social psychology that people tend to underestimate the influence of social pressure on others' public actions (e.g., Miller et al., 1974). In particular, we assume that this underestimation is a consequence of using a heuristic to simplify a decision problem under uncertainty. What is more, our assumption that players' heuristics do not correct themselves over time draws upon empirical evidence in economics that a lack of communication can sustain incorrect judgment about others' beliefs (Bursztyn et al., 2020). We adapt the modelling of preferences in Kuran (1987a,b, 1995) when assuming that players face a trade-off between social approval and personal autonomy: conforming may preclude truthfully revealing one's type. However, our players move sequentially rather than simultaneously to study how heuristic decision-making interacts with an informational asymmetry to trigger an action that differs from one's type.

This essay contributes several insights for policy. In our analysis of individual behaviour, we show that a policy maker can boost a credulous player to scepticism if his goal is to increase truth telling. In contrast, a boost to rationality does not guarantee that more players truthfully reveal their type. An intervention that lowers the pressure to conform in the presence of heuristic decision-making may be more effective. In our analysis of aggregate behaviour, we show that a boost to rationality is not only necessary to incentivise players to take the action matching the state irrespective of their types, but also preferable given a utilitarian objective. Finally, making the state of the world common knowledge can be a viable alternative when a policy maker prefers a player to tell the truth if her type matches the state and to falsify if it does not.

We highlight the implications of our results and the welfare impact of heuristic decision-making in two applications that vary the nature of players' actions and types – organisational decision-making and social engineering. A board or committee, for example, may reach a consensus at all costs without critical appraisal of alternatives given their members' desire for conformity. In this situation, a policy maker may want to boost credulous players to scepticism or reduce conformity to limit the erroneous "infer[ence] that the identical actions of the self and others reflect different internal states" (Miller and McFarland, 1987, p.298), thereby improving the board's collective decision-making. In other situations, rationality, an informative environment, or even common knowledge of the state will be optimal for positive social change: locking in an action early on may be desirable to maximise social welfare. This may allow a policy maker to incentivise a re-thinking on important issues, such as climate change in the community.

Each of my essays is motivated by the observation of a real-world phenomenon that warrants a better understanding through economic modelling. The models they inspire do not only provide testable predictions and insights for policy, but connect in two ways. First, all include behavioural elements through biases, heuristics, or an enlarged set of preferences. Second, all deploy longstanding concepts from microeconomic theory and information economics, including verifiable disclosure, signalling, and private types. Through this connection, I contribute to a way of doing economics that is relevant to society.

Chapter 2

First Impression Biases in the Performing Arts: Taste-Based Discrimination and the Value of Blind Auditioning*

2.1 Introduction

> *"I've been in auditions without screens, and I can assure you that I was prejudiced. I began to listen with my eyes, and there is no way that your eyes don't affect your judgment. The only true way to listen is with your ears and your heart."*
> – Julie Landsman in Gladwell, *Blink: The Power of Thinking Without Thinking*

Auditions are common practice in the performing arts to hire the best candidate. The evaluator, who makes the hiring decision, chooses upfront between two audition forms. If *informed*, the evaluator requires applicants to submit information prior to the audition. This often takes the form of a CV which contains valuable information about ability. One proxy for ability is the school an applicant went to, such as Juilliard, which leads the evaluator's belief about her ability to increase. A CV, however, also contains irrelevant information about gender or race.[1] This may result in *first impression biases* which discount more relevant information obtained at a later stage (Thorngate et al., 2010). If *blind*, the evaluator receives no information and, because the evaluator cannot see the applicant during her performance, the hiring decision is based purely on sound. Given the audition

*Published (excl. additional materials in Section A.1) in the *Journal of Cultural Economics* on 8 September 2021.

[1] In countries like France, Belgium or Germany, it is common to include a photograph that likely reveals gender or race. Even in countries like the UK where it is advised against the inclusion of a photograph or personal details like birthplace due to equal opportunity legislation, the first and last name may reveal gender or race; see Bertrand and Mullainathan (2004) for an experiment that randomly assigns African-American- or White-sounding names to CVs.

form, the applicant decides whether to participate and how much effort to invest in preparing a performance. Importantly, this sequential decision-making of the potentially biased evaluator and the applicant allows the audition form to affect effort incentives.

In this paper, I analyse the repercussions of the evaluator's bias against a specific group of applicants. I examine whether, and under which conditions, it can be rational for the evaluator to commit to ignore supplementary information about an applicant to avoid misleading first impressions. I ask whether the move to a blind audition can lead to better candidates being hired; in essence, can a blind audition provide better effort incentives exclusively for highly able applicants?

My main contribution lies in providing a theory for Goldin and Rouse's (2000) finding that the use of blind auditions in U.S. orchestras has lead to a severalfold increase in the probability that a female will be hired. Furthermore, I address the two polarised views on the use of blind auditions to hire the best candidate. Proponents argue that blind auditions have helped evaluators avoid misleading first impressions (Gladwell, 2005).[2] In fact, blind auditions have often been marketed as a fairer form of audition and equal opportunity policy (Karl Schiebler in The Economist, 1996).[3] Opponents fear that, while avoiding bias, the evaluator loses valuable information that can screen applicants by ability (Blasko in Makoff-Clark, 2019). Consequently, top talent may be lost in an increasingly large applicant pool and standards of fairness may even deteriorate (Holland, 1981). My second contribution is to show that the two polarised views are not mutually exclusive. I argue that a blind audition comes with a fundamental trade-off: it grants impartiality to the evaluator but may prohibit the screening of applicants by ability.

I develop a game-theoretic framework in which the evaluator is biased against females and his bias is common knowledge. The evaluator commits to a blind or an informed audition given an exogenously determined wage. He always fills the position with an applicant or a random outside option. His revenue is determined by performance quality and ability whereas the applicant cares only about being hired. In the benchmark model, effort maps one-to-one into performance quality. In an extension, I allow for moral hazard by considering a setting where low effort can result in low or high performance quality. I focus on a gender bias and frequently use the context of an orchestra to build intuition. This focus also emphasises that artistic auditions are a key application as the underrepresentation of

[2] Gladwell (2005) bases his argument on the prejudices that women have encountered in the past when auditioning for orchestra positions. Two examples are not possessing the physical strength for instruments traditionally considered "male", or an inferior attitude and resilience compared to males.

[3] John de Mol, the creator of the singing competition "The Voice", for example, argues that the TV show's blind auditions allow the judges to focus on quality alone and identify genuine singing talent (Universal Music Group, 2011).

women in virtually all the arts provides evidence for some degree of discrimination (Hellmanzik, 2020).[4] However, my framework pertains to a broader range of contexts in which the evaluator requires the applicant to prepare a project for review, and possibly submit additional information susceptible to bias, before deciding upon acceptance. In particular, technology companies like 'GapJumpers' (Cain Miller, 2016) or 'Applied' (Glazebrook, 2019) allow firms to implement blind hiring through software that withholds personal information.[5] Academia may be another application if blind auditioning can help to address the lower success rates of female applicants for funding schemes such as ERC grants (Vernos, 2013).[6]

My main result is that there exists a threshold bias above which the evaluator prefers a blind audition to provide *targeted effort incentives*: high effort incentives for highly able applicants while low-ability applicants do not participate. Consequently, committing to no information can be beneficial for the evaluator. Moreover, above the threshold bias, equal opportunity and targeted effort incentives are complementary objectives: there is no trade-off between implementing blind auditions as a policy to counteract first impression biases and efficiency due to adverse effects on targeted effort provision. If anything, the implementation of equal opportunity is effort-enhancing for high-ability applicants above this threshold. Perhaps surprisingly, the preferences of a highly biased evaluator align with the preferences of high-ability females: a blind audition provides targeted effort incentives and, as a side effect, plays out in favour of high-ability females, as they prefer to conceal their gender. My model sheds light on the forces underlying the empirical findings on the use of blind auditions and shows that the gains from being able to conceal one's gender need not be uniform. For a low bias, the gains from moving to a blind audition accrue to low-ability females. If the bias is high, the gains accrue to high-ability females.

I show in an extension that a blind audition cannot provide targeted effort incentives if performance uncertainty exceeds some threshold. Consequently, for a highly biased evaluator, a blind audition is no longer guaranteed to be more profitable than an informed audition. Moreover, there is the sizeable risk of market failure: for certain parameter combinations, neither a blind nor an informed audition is profitable and the evaluator might not want to hold an audition at all. Interventions that ensure the applicant pool to be of sufficiently high ability under performance uncertainty guarantee the profitability of a blind

[4]See Chapter 5 in Hellmanzik (2020) for an excellent case study of discrimination in artistic labour markets and a discussion of the classic work by Goldin and Rouse (2000).

[5]With a commonly known gender bias, such intermediaries may also serve as a commitment device to blind hiring.

[6]The act of withholding applicants' names may already have a significant effect; see, for example, Bertrand and Mullainathan's (2004) experiment. For an example of name-blind hiring, see the UK Civil Cervice (Manzoni, 2015).

audition when the evaluator is highly biased and avoid market failure. Finally, I show that a policy maker may want to allow a moderately biased evaluator to request a gender-blind CV prior to a blind audition.

2.1.1 Related Literature

My paper relates to several strands of literature. I draw upon the literature on cognitive biases: I argue that the evaluator in an informed audition is likely subject to different types of biases which may be summarised in a reduced-form parameter. One potential driver is intergroup bias: the evaluator may judge members of his own group more favourably (Hewstone et al., 2002; Anwar, 2012; Bagues and Perez-Villadoniga, 2012). In orchestras, gender associations with instruments may drive bias (Abeles, 2009; Stronsick et al., 2018; Sergeant and Himonides, 2019). A third potential driver is due to interaction effects of gender and race: in experimental judging sessions, videotaped males tend to score lowest among black performers; females lowest among whites (Elliott, 1995). Extraneous factors warrant the modelling of bias in performance evaluation. This can be as subtle as the strength of the applicant's handshake at the audition if this opening ritual influences the evaluator's belief about her character strength.[7] Other examples are attractiveness or accent (Purkiss et al., 2006).[8]

I draw upon contract theory by assuming verifiable disclosure (Milgrom, 1981; Grossman, 1981): in an informed audition the applicant cannot misrepresent her gender nor ability on her CV. Conversely, the evaluator cannot commit to cherry-pick relevant information. Effort incentives are, therefore, influenced by the evaluator's bias via his hiring rule. This feature is similar to implicit incentives arising in a dynamic setting (Meyer and Vickers, 1997). The evaluator has a menu of audition forms available to hire the best candidate and, given his commonly known bias, needs to choose the optimal form (Laffont and Tirole, 1986).[9] Moreover, the blind audition is akin to a signalling model à la Spence (1973) as the applicant's effort decision may be informative about the ability dimension of her type while gender cannot be signalled.

[7]In particular, "[b]ecause men tend to have greater upper body strength than do women, a male's handshake is likely to be firmer than a female's handshake". As a result, "if the [evaluator] shakes a woman's hand, it is more likely the shake will fail the [evaluator]'s character test", leading him to discount more relevant information obtained from the female's subsequent performance (Thorngate et al., 2010, p.51).

[8]There is evidence for a vision heuristic: evaluators consistently report to value sound as central in performance. However, they select different winners if they have visual information about the applicant (Tsay, 2013).

[9]In Laffont and Tirole (1986), the evaluator would be the agent. The principal would be a regulator in charge of what audition forms are available to the evaluator; for example, the admission of gender-blind CVs (Section 2.9).

My paper builds on literature in the economics of discrimination. In models of statistical discrimination (Phelps, 1972; Arrow, 1973), the evaluator has to believe that women are, on average, less productive than men to use gender in the hiring decision as a proxy for unobservable ability. In taste-based models (Becker, 1971), the evaluator instead has a preference for hiring members of the dominant group; say, men. Also an evaluator's self-image bias (Siniscalchi and Veronesi, 2020) or type-based mentoring (Athey et al., 2000) can perpetuate gender imbalance. I draw on experimental evidence that discrimination widens labour market differences between women and men in male occupations (Riach and Rich, 2006), and that women can suffer from discrimination based on their physical appearance (Alaref et al., 2020).

Blind auditions are studied by Taylor and Yildirim (2011), focusing on the effort and selection effects when the evaluator observes a signal of performance quality. Their model can explain the association of blind auditions with an increased probability of hiring females through a form of statistical discrimination, as a blind audition forces the evaluator to ignore the applicant's ability. Similar to Taylor and Yildirim (2011), I study the conditions under which a blind audition provides better effort incentives. However, I allow for taste-based discrimination of the evaluator so that a blind audition forces him to ignore the applicant's ability *and* gender. In my model, females and males are equally productive for the same level of ability. This assumption stresses that prejudices against female musicians (Seltzer, 1989), social attitudes (Allmendinger and Hackman, 1995) and self-identity (Starr, 1974) are key in explaining the findings of Goldin and Rouse (2000).

When the evaluator in Taylor and Yildirim (2011) can commit to an acceptance threshold prior to the audition and the observed quality signal is verifiable, he is forced to reject performances that he knows to be of high quality with positive probability to provide effort incentives. Such a trade-off at the high end of the quality spectrum does not arise in my informed audition; rather, applicants with low ability may be screened as they anticipate a lower probability of being hired. Moreover, in Taylor and Yildirim's (2011) blind audition, the evaluator uses a uniform acceptance threshold. My blind audition, in contrast, allows for signalling concerns: a uniform hiring probability only arises when applicants pool on high effort.

In Taylor and Yildirim (2011), if the applicant pool is of mainly high ability, the evaluator prefers a blind audition as assessing performance quality is less important than providing effort incentives; if the applicant pool is of predominantly low ability, the evaluator prefers an informed audition as selecting high-quality performances is crucial. Taste-based discrimination, in contrast, allows for richer predictions: if there is little bias and the applicant pool is highly able, my blind audition attracts low-ability applicants rather than providing targeted effort incentives. For a high bias, the evaluator prefers a blind audition irrespective of the applicant pool's ability composition.

In Taylor and Yildirim (2011), if the signal of performance quality is precise, the evaluator prefers a blind audition as the incremental information from observing ability is low; if imprecise, he prefers an informed audition as observing ability has a large incremental informational value. My model, in contrast, makes more nuanced predictions about the effect of performance uncertainty: if there is little bias, informed auditions are preferred irrespective of the degree of performance uncertainty. If the bias is high and performance uncertainty is above a threshold, the ability composition of the applicant pool determines the evaluator's preferences. For a very low prior about ability, it is paramount to include ability in the hiring policy. While blind auditions provide better effort incentives for the highly able, these types are infrequent. For a sufficiently high prior, high effort incentives for the highly able are paramount, making blind auditions preferable.

The remainder of this paper is structured as follows. Section 2.2 sets out preliminaries common to both auditions. In Section 2.3 and Section 2.4, I solve for equilibrium in the informed and blind audition. Section 2.5 details my main results on the evaluator's and applicant's audition preferences. I introduce asymmetric uncertainty in Section 2.6. I compare blind auditions to quotas and subsidies in Section 2.7. I briefly highlight the agency cost in the two audition forms in Section 2.8. I discuss the use of gender-blind CVs in Section 2.9, and conclude with avenues for further research in Section 2.10.

2.2 Preliminaries

In this section, I set out the timing, preferences of players, assumptions and solution concepts.

2.2.1 Timing

Two risk neutral parties, an applicant (she) and an evaluator (he) play a three-stage game.

In the first stage, the evaluator commits to a blind (B) or an informed (I) audition given an exogenously determined wage. If the audition is informed, the evaluator learns the applicant's type $\theta := (\eta, g)$, where $\eta \in \{\eta_L, \eta_H\}$ is ability[10] and $g \in \{m, f\} := \{0, 1\}$ is gender. If the audition is blind, the evaluator does not learn the applicant's type. He only knows the type distribution: the applicant is female or male with equal probability and the

[10] Ability may equate to an artist's creativity (Towse, 2006). See also Duff's (1767) classic work on original genius: a highly able applicant may have the right combination of imagination, judgment and taste to (i) prepare an original performance for the audition, and (ii) reduce the investment needed to attain a given level of mastery.

applicant is of high ability with $\Pr(\eta_H) := p \in (0, 1)$. It is common knowledge that ability and gender are independent.

In the second stage, the applicant moves, knowing her type and whether it is a blind or informed audition. She chooses how much costly effort[11] to invest in preparing a performance, where effort $e \in \{e_L, e_H\}$ can be low or high and effort costs are $c(e)$. There is an outside option for the applicant that yields a zero payoff.

In the third stage, the applicant delivers her performance. It can turn out to be of low or high quality; that is, $q \in \{q_L, q_H\}$. The evaluator makes the hiring decision based on type θ (i.e. ability and gender) and performance quality q if the audition is informed, and based on performance quality q alone if the audition is blind.[12] In either audition form, the evaluator has a random outside option $\bar{U}$, which he can revert to instead of hiring the applicant. I describe the evaluator's outside option and the determination of performance quality in more detail in Section 2.2.3.

2.2.2 Preferences of Players

The evaluator is assumed to be biased against female applicants. This is represented in form of a bias parameter $\beta \in [0, q_L + \eta_L]$. The level of bias β is common knowledge in either audition form. I formalise the evaluator's preferences with a gross utility function consisting of two parts:

$$V(q, \eta, g) = f(q, \eta) - \beta g = q + \eta - \beta g. \tag{2.1}$$

The first part $f(q, \eta) = q + \eta$ is a production function.[13] The evaluator, therefore, prefers to hire an applicant who is of high ability and delivers a high-quality performance in stage 3. In particular, given the applicant's ability η, which the evaluator may learn in stage 1, and her performance quality q, always observed in stage 3, the evaluator earns revenue $f(q, \eta)$ by hiring the applicant. The assumption that performance quality and ability jointly determine revenue may be justified if $f(q, \eta)$ is regarded more broadly as an applicant's marketable

[11] In the performing arts, an applicant's effort includes the time spent studying the set audition repertoire to determine which excerpt is appropriate, the time spent rehearsing the selected excerpt, etc. Without an effort choice, the game would reduce to a decision problem of the evaluator, forbidding the study of incentive effects in light of bias.

[12] An alternative interpretation of the game involves two evaluators: the senior management with a zero bias chooses the audition form in stage 1; the evaluator with a possibly non-zero bias chooses whom to hire in stage 3. The senior management may have learned the bias only after assigning the evaluator to the task.

[13] Under a multiplicative form of the production function, $f(q, \eta) = q\eta$, most of the results remain qualitatively unchanged. If the applicant's ability did not enter the production function directly, there would not be a motive to signal high ability in the blind audition, providing targeted effort incentives at high biases in Section 2.4. Under $f(q) = q$, the evaluator would, however, not be indifferent between audition forms: for a moderate bias, he would prefer a blind audition for targeted effort incentives; for biases above a certain threshold, he would not hold an audition at all. The proofs are available on request.

talent, consisting of an endogenous component q and an exogenous component η.[14] The inclusion of ability allows for revenue-generating reputation effects from hiring highly able star soloists[15] that have been shown to attract fans (Hart, 1973, p.390) and additional single-ticket buyers (Kamakura and Schimmel, 2013). Furthermore, the functional form of $f(q, \eta)$ implies that males and females are perfect substitutes in production: gender does not enter. The second part βg captures the evaluator's bias against female applicants.[16]

The evaluator faces four types of applicants in this game, $\Theta = \{(\eta_L, m), (\eta_H, m), (\eta_L, f), (\eta_H, f)\}$. The applicant, whether of low or high ability, female or male, wants to be hired to receive a wage w_θ net of effort costs. In particular, by choosing $e \in \{e_L, e_H\}$ she maximises her expected utility U, taking into account how her effort choice influences the probability of being hired via its effect on performance quality:

$$U(e, \eta, g) = \Pr(\text{hired}|\cdot) w_\theta - \frac{e^2}{2\eta}. \tag{2.2}$$

In an informed audition, the hiring probability depends on performance quality, gender and ability; that is, $\Pr(\text{hired}|q, \eta, g)$. In a blind audition, in contrast, the hiring probability never directly depends on gender or ability; that is, $\Pr(\text{hired}|q)$. The applicant's effort costs are given by $c(e) = \frac{e^2}{2\eta}$. Thus, ability determines the marginal cost of exerting high rather than low effort, with high ability corresponding to lower marginal costs. The applicant receives a zero payoff if she is not hired and her outside option gives zero utility.

2.2.3 Assumptions and Solution Concept

I assume that the evaluator's revenue from a high level of ability and performance quality is twice as large as from a low level, and that ability and performance quality are equally valuable to the evaluator.

Assumption 1. $\eta_L = e_L = q_L = 1$ *and* $\eta_H = e_H = q_H = 2$.

Given that the bias is bounded such that a maximally biased evaluator's gross utility from hiring a low-ability female with a low-quality performance is zero, I can normalise the parameter levels in Assumption 1 without loss of generality: my model would deliver identical results if $\eta_L = e_L = q_L = x$ and $\eta_H = e_H = q_H = 2x$, where $x > 0$. Assumption 1 keeps my

[14] See Tsay and Banaji (2011) for the two sources of talent and how they are perceived by expert decision-makers: "naturals" are characterised by early evidence of high innate ability; "strivers" by high motivation and perseverance. Their substitutability is in line with Galenson's (2006; 2009) finding that there are two equally successful types of artists: experimental artists improve their work by trial and error; conceptual artists work with deductive certainty.

[15] See also the economics literature on the phenomenon of superstars (Adler, 1985; Rosen, 1981).

[16] This form of the utility function is inspired by (Becker, 1971, chapter 3, p.39): an employer has a taste parameter d ("coefficient of discrimination") so that the wage of a member of the minority group N is effectively $w_N + d$.

model tractable but rules out general conclusions on the effect of blind auditions: my model provides insights when the relationship between ability, effort and performance quality is balanced from the evaluator's perspective.[17]

I assume that the evaluator always fills the position by either hiring the applicant at a wage w_θ or a uniformly distributed outside option at a wage $w_{\bar{U}}$. An unfilled orchestra position during the playing season, for example, would be prohibitively costly. As concert schedules are planned far in advance, ticket sales tend to start months before the performances and orchestral parts need to be acquired and allocated prior to rehearsals (Towse, 2010, chap. 8), the evaluator is unlikely to leave such a position unfilled.

Assumption 2. $\bar{U} \sim \mathcal{U}[2-\beta, 4]$.

The value of $\bar{U}$ is unknown to the evaluator when making the hiring decision. The bounds on the outside option in Assumption 2 may be justified if $\bar{U}$ is thought of as the possibility to engage a substitute from an agency. This substitute generates gross utility of at least $2-\beta$ if she is of type $\theta = (\eta_L, f)$ and provides a low performance quality. The substitute generates at most 4 if he is of type $\theta = (\eta_H, m)$ and provides a high performance quality. Because the evaluator cannot contract upon the substitute's type and performance quality with the agency, $\bar{U}$ is essentially a gamble that may turn out to be better or worse than the applicant considered in stage 3 of the game. Given this interpretation, a greater bias against female applicants naturally reduces the minimum value of the outside option. In fact, the outside option may turn out to be of no value to a maximally biased evaluator. Continuity is assumed for tractability; however, it implies a discrepancy between the number of types of the outside option and the applicant in the model.

I assume that the evaluator pays wage w whether he hires an applicant of type θ or the outside option, and that the wage received by the applicant and the outside option is exogenously determined by the outside option's expected value $\mathbb{E}[\bar{U}]$.

Assumption 3. $w := w_\theta = w_{\bar{U}}$ *for all* $\theta \in \Theta$ *and* $w = \mathbb{E}[\bar{U}]$.

[17] A more general framework where high talent (η_H and q_H) can be more or less than twice as valuable allows for qualitatively different equilibria: a low value coupled with a high bias, for example, will discourage all females from participating in an informed audition while attracting all males with no incentive to exert high effort given their possibility to identify themselves as male. A blind audition in this setting will *not* provide targeted effort incentives either but may at least achieve gender balance in the applicant pool that participates in the audition. A more general framework where a given level of ability and performance quality need not be equally valuable eliminates the symmetry $\Pr(h|q_L, \eta_H, \cdot) = \Pr(h|q_H, \eta_L, \cdot)$ in an informed audition: when performance quality becomes slightly more valuable, low-ability males are attracted at even lower biases but the highly able exert high effort at greater biases. In this setting, the evaluator's bias must be more severe before a blind outperforms an informed audition, and at extreme biases the blind audition may increase the applicant pool that participates rather than reducing it to the highly able.

Assumption 3 is in line with orchestras' practice to specify a base pay according to which positions are remunerated, standardising wages across audition forms for applicants and the outside option.[18] Furthermore, fixing the wage at the expected value of the outside option captures that trade unions tend to be integrated in the setting of the base pay to performers (Towse, 2010): the base pay reflects the market value of the outside option ex-ante and is the evaluator's maximum willingness-to-pay when negotiating with unions. Given Assumption 3, I can abstract from wages in the evaluator's utility function (2.1). Wage costs are sunk and do not affect his hiring decision in stage 3. Later, I will refer to the evaluator's utility net of wage costs as Π.

In the benchmark model, I focus on the interaction of the strength of the evaluator's bias and the information structure of the audition.

Assumption 4. $\Pr(q_i|e_i) = 1$ *for* $i \in \{L, H\}$.

Assumption 4 abstracts from any moral hazard considerations: effort maps one-to-one into performance quality, so can be perfectly inferred by the evaluator. In Section 2.6, I relax this assumption and allow for moral hazard by considering a setting where low effort can result in low or high performance quality. Given Assumption 4, an informed audition is a game of complete information, and the solution concept is backward induction. A blind audition is a game of incomplete information, and the solution concept is perfect Bayesian equilibrium, refined by the D1-Criterion when necessary.

2.3 Informed Audition

If the evaluator has committed to an informed audition, stage 2 and 3 are a game of complete information with four proper subgames, one for each applicant type. I, therefore, solve the game by first considering the evaluator's hiring decision in stage 3. Using this decision, I then determine the applicant's optimal effort choice in stage 2.

2.3.1 Hiring Decision of Evaluator

The evaluator compares the utility from hiring the applicant to the value from hiring the outside option. Therefore, he hires the applicant in stage 3 with probability

$$\Pr(\text{hired}|q, \eta, g) := \Pr(\bar{U} \leqslant V(q, \eta, g)) = \frac{V(q, \eta, g) - (2 - \beta)}{4 - (2 - \beta)} = \frac{q + \eta + \beta(1 - g) - 2}{2 + \beta}. \tag{2.3}$$

[18] In the UK, for example, employment agreements are negotiated by the Musicians' Union (n.d.) and applicable to members employed with major orchestras. The assumption also eliminates the screening aspect that wage setting would otherwise have in stage 1. In particular, endogenising wages would make a blind audition preferable even for an unbiased evaluator due to significant wage savings. Furthermore, the evaluator would then be best off setting $w = 0$, always hiring the outside option because $\bar{U}$ does not respond to incentives in this model.

Intuitively, the evaluator considers how likely it is that the substitute from an agency will deliver a gross utility less than $V(q, \eta, g)$ from hiring the applicant, and makes this probability his hiring rule.

Given Assumption 4 and the hiring rule (2.3), there are eight possible hiring probabilities in an informed audition:

$$\Pr(h|q_L, \eta_L, f) = 0, \qquad \Pr(h|q_L, \eta_L, m) = \frac{\beta}{2+\beta},$$

$$\Pr(h|q_H, \eta_L, f) = \Pr(h|q_L, \eta_H, f) = \frac{1}{2+\beta}, \qquad \Pr(h|q_H, \eta_L, m) = \Pr(h|q_L, \eta_H, m) = \frac{1+\beta}{2+\beta},$$

$$\Pr(h|q_H, \eta_H, f) = \frac{2}{2+\beta}, \qquad \Pr(h|q_H, \eta_H, m) = 1.$$

Importantly, the evaluator never hires a low-ability-low-effort female in an informed audition: by reverting to the outside option, he can never do worse but may be lucky to engage a substitute of higher ability, higher effort or different gender. On the other hand, the evaluator always hires a high-ability-high-effort male: by reverting to the outside option, the evaluator cannot do any better. The remaining hiring probabilities depend on the strength of the evaluator's bias. Intuitively, for females, the hiring probabilities are decreasing in β. For males, those are increasing in β.

2.3.2 Effort Decision of Applicant

Given the hiring probabilities, the applicant chooses effort to maximise her expected utility (2.2). To determine her optimal effort response as a function of the bias, I compare the participation and incentive constraint for each applicant type.

Lemma 1. *The effort responses of the four applicant types partition the evaluator's bias $\beta \in [0, 2]$ into three regions: (i) For a low bias, $\beta \in \beta_I^L := [0, \frac{5-\sqrt{17}}{2})$, only high-ability applicants participate and exert high effort.*[19] *(ii) For a moderate bias, $\beta \in \beta_I^M := [\frac{5-\sqrt{17}}{2}, \frac{6}{5}]$, low-ability males also participate and exert low effort. (iii) For a high bias, $\beta \in \beta_I^H := (\frac{6}{5}, 2]$, all applicants except low-ability females participate and exert low effort.*

Proof. Detailed proofs of all results are in Section B.1. □

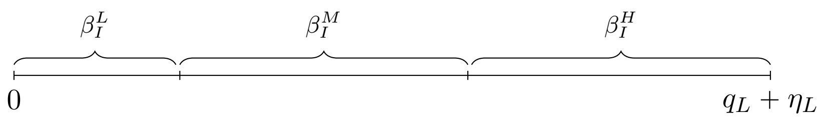


Figure 2.1: Partition of the evaluator's bias in an informed audition.

[19] Note that $\frac{5-\sqrt{17}}{2} \approx 0.44$.

The partition of the evaluator's bias into three regions is illustrated in Figure 2.1. The low and moderate bias cover roughly 21.9 and 38.1 percent of the parameter space, respectively. The high bias is the largest interval, covering 40 percent.

2.3.3 Expected Utility of Evaluator

The evaluator's prior is that the applicant is equally likely to be a female or male when committing to the audition form. Furthermore, he knows that the applicant is of high ability with probability $p \in (0,1)$, and that ability is independent of gender. Hence, all four proper subgames are reached with nonzero probability and are, thus, relevant when deriving the evaluator's expected net utility for different values of β.

Proposition 1. *(i) Suppose $\beta \in \beta_I^L$ where β_I^L is defined as in Lemma 1(i). Then,*

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^L] := \mathbb{E}[V|\beta \in \beta_I^L] - w = \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right) + \left(1+\frac{\beta}{2}\right)\right] > 0. \tag{2.4}$$

(ii) Suppose $\beta \in \beta_I^M$ where β_I^M is defined as in Lemma 1(ii). Then,

$$\begin{aligned}\mathbb{E}[\Pi_I|\beta \in \beta_I^M] &:= \mathbb{E}[V|\beta \in \beta_I^M] - w,\\ &= \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right) + \left(1+\frac{\beta}{2}\right)\right] \gtrless 0.\end{aligned} \tag{2.5}$$

(iii) Suppose $\beta \in \beta_I^H$ where β_I^H is defined as in Lemma 1(iii). Then,

$$\begin{aligned}\mathbb{E}[\Pi_I|\beta \in \beta_I^H] &:= \mathbb{E}[V|\beta \in \beta_I^H] - w,\\ &= \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{1}{2+\beta}\left(-\frac{\beta}{2}\right) + \frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)\right] \gtrless 0.\end{aligned} \tag{2.6}$$

If the evaluator's bias is low, he breaks even in subgame (η_L, f) and (η_L, m), and expects positive net utility in subgame (η_H, f) and (η_H, m). Intuitively, when the evaluator is almost impartial, requiring a CV acts as an effective means to screen applicants by ability and to provide targeted effort incentives: high effort incentives exclusively for high-ability applicants while low-ability applicants do not participate. If the bias is moderate, the evaluator instead expects negative net utility in subgame (η_L, m) as low-ability males exert low effort rather than dropping out. Low-ability males essentially exploit the evaluator's bias: when they know the bias to be moderate, the probability of being hired conditional on exerting low effort and identifying themselves as male is sufficiently high to make the cost of low effort worthwhile. If the evaluator is highly biased, he also expects negative net utility in subgame (η_H, f). Intuitively, when high-ability females know the evaluator's bias to be substantial, the marginal cost from putting in high rather than low effort outweighs the marginal benefit from being hired more often. They are deterred from exerting high effort at such a high bias.[20] Moreover, high-ability males rest on their laurels and have little incentive to exert high effort. Their benefit from exploiting the bias and being able to identify themselves as male makes the increase in cost from exerting high rather than low effort not worthwhile.

Proposition 1 highlights the role of ability. For a low bias, the evaluator's expected net utility (2.4) is increasing in the prior that the applicant is of high ability. An increasingly able applicant pool makes it more likely that the evaluator faces the profitable subgame (η_H, f) or (η_H, m) relative to the subgames in which he breaks even with the outside option. For a moderate bias, the increase in the evaluator's expected net utility (2.5) is more pronounced: as subgame (η_L, m) becomes less likely, the evaluator is able to avoid the losses from potentially hiring low-ability males who participate in the audition. For a high bias, the increase in the evaluator's expected net utility (2.6) is more attenuated: as p increases, the evaluator faces both the unprofitable subgame (η_H, f) and the profitable subgame (η_H, m) more often.

2.4 Blind Audition

In a blind audition, the evaluator does not learn the applicant's type θ in stage 1. He only observes her performance quality q in stage 3. Thus, there are no proper subgames but two information sets: q_L and q_H. Each information set contains four nodes when the evaluator has to make the hiring decision in stage 3. Because these information sets are non-singletons, stage 2 and 3 in a blind audition are a game of incomplete information. Therefore, I specify beliefs for the evaluator: a probability distribution over the nodes in both q_L and q_H. Intuitively, this probability distribution constitutes the evaluator's revised beliefs how likely it is, upon hearing a low- or high-quality performance, that he is facing a particular applicant type behind the curtain. In what follows, I focus on the case when the bias is either sufficiently low to induce all applicants to exert high effort or sufficiently high to induce only the highly able to exert high effort.[21] I refer to the former case as pooling and the latter case as separating.

[20] The mechanism is similar to stereotype threats (Günther et al., 2010).

[21] I do not consider applicants pooling on low effort as my focus is on whether, and under which conditions, blind auditions can be superior. To see why this can never obtain in a low-effort pooling equilibrium, compare it to the worst-case scenario in an informed audition; that is, for $\beta \in \beta_I^H$, all applicants except low-ability females exert low effort. This scenario gives the evaluator still a higher payoff because attracting low-ability-low-effort females is never profitable. The evaluator could have broken even with the outside option and, therefore, is worse off.

2.4.1 Pooling Equilibrium

Suppose the evaluator's bias is sufficiently low to induce all applicants to exert high effort. However, when observing a low performance quality off the equilibrium path, the evaluator believes the applicant to be of low ability. Upon observing a low-quality performance, the evaluator can, therefore, infer the applicant's ability from her action and expects gross utility

$$\mathbb{E}[V|q_L] = q_L + 1 - \frac{\beta}{2} = 2 - \frac{\beta}{2}$$

from hiring the applicant.[22] Upon observing a high-quality performance, the evaluator cannot infer the applicant's ability and holds a belief that is equal to his prior. He expects gross utility

$$\mathbb{E}[V|q_H] = q_H + \mathbb{E}[\eta|q_H] - \frac{\beta}{2} = 2 + [(1-p) + 2p] - \frac{\beta}{2} = 3 + p - \frac{\beta}{2}$$

from hiring the applicant. The evaluator compares the expected utility from hiring the applicant to the value from hiring the outside option. Therefore, he hires the applicant at each information set with probability

$$\Pr(\text{hired}|q) := \Pr(\bar{U} \leqslant \mathbb{E}[V|q]) = \frac{\mathbb{E}[V|q] - (2-\beta)}{4-(2-\beta)} = \frac{q + \mathbb{E}[\eta|q] - \frac{\beta}{2} - (2-\beta)}{2+\beta}. \tag{2.7}$$

Given Assumption 4 and the hiring rule (2.7), there are two possible hiring probabilities for both female and male applicants:

$$\Pr(h|q_L) = \Pr\left(\bar{U} \leqslant 2 - \frac{\beta}{2}\right) = \frac{\beta}{4+2\beta},$$
$$\Pr(h|q_H) = \Pr\left(\bar{U} \leqslant 3 + p - \frac{\beta}{2}\right) = \frac{2(1+p)+\beta}{4+2\beta}.$$

Compared to an informed audition, female and male applicants now face the same hiring probabilities as the evaluator cannot bias gender. Similarly, high- and low-ability applicants face the same hiring probabilities as they cannot identify themselves in a pooling blind audition. Furthermore, an increasingly skilled applicant pool raises the evaluator's expected revenue from hiring at q_H. Hence, he finds it optimal to hire more often at this information set.

[22] The evaluator cannot infer gender in a pooling or separating blind audition because, with η fixed, females and males have the same payoff structure. In effect, the evaluator now faces only two applicant types, η_L and η_H.

2.4.1.1 Effort Decision of Applicant

Given the two hiring probabilities, the applicant chooses effort to maximise her expected utility (2.2). I now show that the prior determines whether a high-effort pooling equilibrium can be supported, and that p places an upper bound on the evaluator's bias in any such pooling equilibrium. I define $\beta(p) := \{\beta : U(e_H, \eta_L) = 0\}$.

Lemma 2. *(i) For $p < \frac{1}{3}$, no high-effort pooling equilibrium exists. (ii) For $p = \frac{1}{3}$, a high-effort pooling equilibrium exists if the evaluator is unbiased. (iii) For $\frac{1}{3} < p < 1$, a high-effort pooling equilibrium exists if $\beta \in \beta_B^L := [0, \beta(p)]$, where $\beta'(p) > 0$. As $p \uparrow 1$, the upper bound on the pooling equilibrium $\beta(p)$ approaches the infimum beyond which a separating equilibrium exists (Lemma 3).*

With the observation that it is always optimal for high-ability applicants to exert high effort if it is optimal for low-ability applicants to exert high effort, pooling on high effort is possible if the prior that the applicant is of high ability is sufficiently large and the evaluator's bias is not too high. Furthermore, an increasingly able applicant pool makes it easier for the fewer low-ability to hide behind the more-and-more high-ability applicants.

2.4.1.2 Expected Utility of Evaluator

While the evaluator does not learn the applicant's type, the applicant's payoff structure is common knowledge. In particular, for the evaluator's q_H-belief that all applicants exert high effort to be consistent with the applicants' effort decision, he needs to have a sufficiently low bias.

Proposition 2. *Suppose $\beta \in \beta_B^L$ with $\frac{1}{3} \leqslant p < 1$ where β_B^L is defined as in Lemma 2. Then,*

$$\mathbb{E}[\Pi_B | \beta \in \beta_B^L] := \mathbb{E}[V | \beta \in \beta_B^L] - w = p\frac{2(1+p)+\beta}{4+2\beta} > 0. \tag{2.8}$$

The evaluator's expected net utility (2.8) is increasing in the prior that the applicant is of high ability because, ceteris paribus, an increasingly skilled applicant pool raises the evaluator's expected revenue from hiring in a pooling blind audition. Intuitively, the revised belief of the evaluator upon observing a high-quality performance takes into account that it is ex-ante more likely to face a high-ability female or a high-ability male behind the curtain.

2.4.2 Separating Equilibrium

Suppose the evaluator's bias is sufficiently high that only high-ability applicants exert high effort. As before, the evaluator believes the applicant to be of low ability when observing a low performance quality. Therefore, the evaluator has degenerate posterior beliefs at both information sets. Upon observing a low-quality performance, he holds the belief that the applicant is of low ability with probability one, and male or female with equal probability. Upon observing a high-quality performance, he believes the applicant to be of high ability with probability one, and male or female with equal probability. At q_L, the evaluator's expected gross utility $\mathbb{E}[V|q_L] = 2 - \frac{\beta}{2}$ from hiring the applicant is unchanged. At q_H, the evaluator can now infer the applicant's ability from her action and expects gross utility

$$\mathbb{E}[V|q_H] = q_H + 2 - \frac{\beta}{2} = 4 - \frac{\beta}{2}$$

from hiring the applicant. As before, the evaluator compares the expected utility from hiring the applicant to the value from hiring the outside option. Given Assumption 4 and the hiring rule (2.7), there are two possible hiring probabilities for both female and male applicants:

$$\Pr(h|q_L) = \Pr\left(\bar{U} \leqslant 2 - \frac{\beta}{2}\right) = \frac{\beta}{4+2\beta},$$
$$\Pr(h|q_H) = \Pr\left(\bar{U} \leqslant 4 - \frac{\beta}{2}\right) = \frac{4+\beta}{4+2\beta}.$$

2.4.2.1 Effort Decision of Applicant

Given the two hiring probabilities, the applicant chooses effort to maximise her expected utility (2.2).

Lemma 3. *For $\beta \in \beta_B^H := (\sqrt{17} - 3, 2)$, a separating equilibrium exists in which only high-ability applicants exert high effort and low-ability applicants do not participate.*

Separation is possible if the bias is sufficiently high so that it is too costly for low-ability applicants to participate. Moreover, for $\beta < 2$, the outside option dominates low effort for low-ability applicants. For $\beta \leqslant 2$, high effort dominates low effort for high-ability applicants.

2.4.2.2 Expected Utility of Evaluator

For the evaluator's q_H-belief that only high-ability applicants exert high effort to be consistent with the applicants' effort decision, he needs to have a sufficiently high bias. In this case, low-ability applicants do not participate and q_L is not reached. Therefore, the evaluator's belief at this information set is not determined by the applicant's equilibrium play. I provide a robustness check in Section A.1.1: two refinements based on either deviation payoffs or the D1-Criterion show that the evaluator's belief that the applicant is of low ability when observing a deviation to q_L constitutes a reasonable restriction.

Proposition 3. *Suppose* $\beta \in \beta_B^H$ *where* β_B^H *is defined as in Lemma 3. Then,*

$$\mathbb{E}[\Pi_B|\beta \in \beta_B^H] := \mathbb{E}[V|\beta \in \beta_B^H] - w = p\frac{4+\beta}{4+2\beta} > 0. \tag{2.9}$$

As before, the evaluator's expected net utility (2.9) is increasing in the prior that the applicant is of high ability. However, the channel through which an increase in the prior leads to an increase in the evaluator's profit is less pronounced. This is because, in a pooling blind audition, an increase in the prior also implied a reduction in the probability that, upon observing a high-quality performance, the evaluator faced a low-ability female or low-ability male behind the curtain.

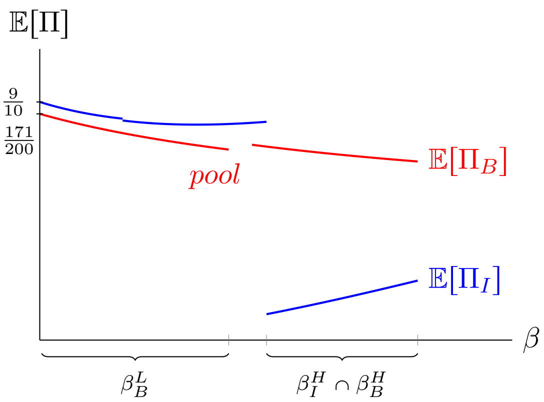
2.5 Comparison of Auditions

Having discussed the evaluator's net utility in both audition forms, I can conclude under which conditions on the bias and the prior the evaluator prefers a blind or informed audition to maximise his expected net utility.

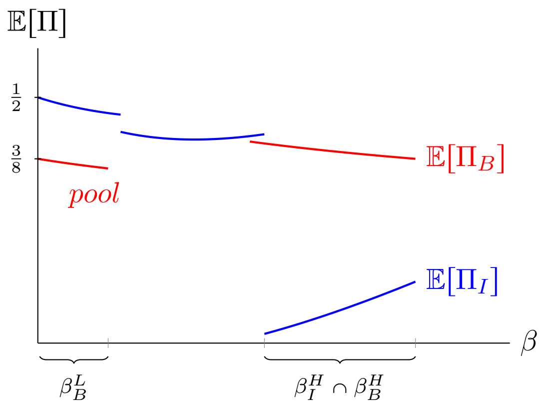
Proposition 4. *(i) For* $\frac{1}{3} \leqslant p < 1$*, if the evaluator's bias against female applicants is low,* $\beta \in \beta_B^L$*, he prefers an informed audition over a pooling blind audition. (ii) For* $0 < p < 1$*, if the evaluator's bias against female applicants is high,* $\beta \in \beta_I^H \cap \beta_B^H$*, he prefers a separating blind audition. (iii) For* $0 < p < \frac{2-\beta}{2}$*, if the evaluator's bias against female applicants is moderate,* $\beta \in \beta_I^M \cap \beta_B^H$*, he prefers a separating blind audition. (iv) For* $\frac{2-\beta}{2} < p < 1$*, if the evaluator's bias against female applicants is moderate,* $\beta \in \beta_I^M \cap \beta_B^H$*, he prefers an informed audition.*

Proposition 4 details my main result on the evaluator's audition preferences: there exists a threshold bias above which the evaluator is better off having no information about the applicant's type. In this case, a blind audition provides targeted effort incentives. In an informed audition, in contrast, the evaluator's bias would distort effort incentives: all applicants except low-ability females would exert low effort. Consequently, committing to no information can be beneficial for the evaluator. This prediction of my model echoes the insight of Gladwell (2005) regarding first impression biases: "by fixing the first impression at the heart of the audition - by judging purely on the basis of ability - orchestras now hire better musicians, and better musicians mean better music" (p.253).

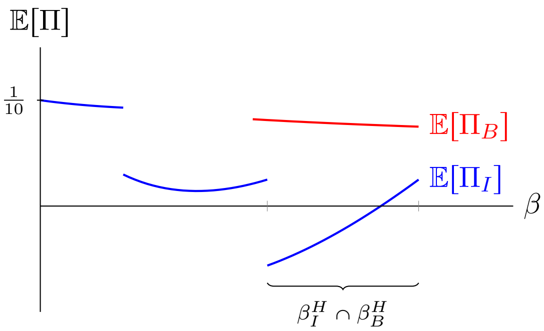
Conversely, the evaluator is better off having more information if his bias against female applicants is low. An informed audition allows him to attract a pool of exclusively high-ability applicants. Moving to a blind audition would also induce low-ability applicants to participate and exert high effort if the evaluator's prior is sufficiently high (Figure 2.2a and Figure 2.2b). Moreover, a blind audition would prevent the evaluator from discriminating between applicants of different characteristics: when all exert high effort, the evaluator has no means of distinguishing them behind the curtain. Specifically, he can no longer hire a high-ability-high-effort male with probability one. However, reverting to the outside option with positive probability when high-ability males exert high effort can only make the evaluator worse off due to the upper bound on $\bar{U}$. This prediction of my model is in line with Holland (1981) observing that “equal opportunity has unloosed such an avalanche of auditionees that standards of selection and fairness are sometimes actually lowered” and that “[a]s a result, top talent is sometimes lost in the shuffle”.



(a) $\Pr(\eta_H) = 0.9$.



(b) $\Pr(\eta_H) = 0.5$.



(c) $\Pr(\eta_H) = 0.1$.

Figure 2.2: Evaluator’s expected net utility in a blind and informed audition.

In Taylor and Yildirim’s (2011) statistical discrimination model, an unbiased evaluator can prefer a blind audition as it provides better effort incentives than an informed audition. In my benchmark without taste-based discrimination ($\beta = 0$), in contrast, the evaluator prefers an informed audition: if at least one in three applicants is expected to be of high ability, a blind audition would lead to an avalanche of auditionees; a blind audition only provides better (i.e. targeted) effort incentives if there is at least some taste-based discrimination, and β exceeds some threshold.

For a small range of moderate biases, $\beta \in \beta_I^M \cap \beta_B^H$, both a blind and informed audition provide high effort incentives for high-ability applicants. However, the evaluator faces the distortion that an informed audition also attracts low-ability males, who exert only low effort. The prior determines the size of the loss from potentially hiring those types and, thus, whether the evaluator may be better off in a separating blind audition. First, for a range of priors from $\frac{2}{5}$ to $\frac{5-\sqrt{17}}{2} \approx 0.44$, the evaluator's preferences over auditions are non-monotonic in his bias (Figure 2.3). Intuitively, this is because equation (2.5) is upward-sloping whereas equation (2.9) is downward-sloping on the interval $\beta_I^M \cap \beta_B^H$. However, if the applicant pool is of predominantly low ability, the evaluator always prefers a blind audition for this range of moderate biases: it is an effective means to provide targeted incentives while reaping the benefits from breaking even with the outside option (Figure 2.2c). The evaluator avoids substantial losses which he would have incurred in an informed audition from hiring low-ability males relatively frequently. As the applicant pool becomes increasingly able, however, this concern becomes less important. In particular, for a sufficiently high prior, the benefit from being able to hire high-ability females and high-ability males with differing probabilities outweighs the loss from potentially hiring a low-ability-low-effort male in an informed audition.

To gain some intuition behind the evaluator's distribution of preferences, suppose that the bias and the prior on high ability are drawn uniformly from their respective supports. In other words, all points in Figure 2.3 are equally likely before the three-stage game is played. Therefore, I can calculate the ex-ante percentage with which a blind audition (red area) is preferred relative to an informed audition (blue area). In so doing, I discount the white areas for which no comparison is possible. The evaluator prefers a blind audition approximately sixty-four percent and an informed audition only approximately thirty-six percent of the time. Therefore, a simple form of heterogeneity in the bias across evaluators and in the ability composition across applicant pools, may rationalise the observed co-existence of blind and informed auditions in the economy. Furthermore, it may rationalise that the majority of U.S. symphony orchestras use a blind audition as an outcome of optimisation (Goldin and Rouse, 2000).

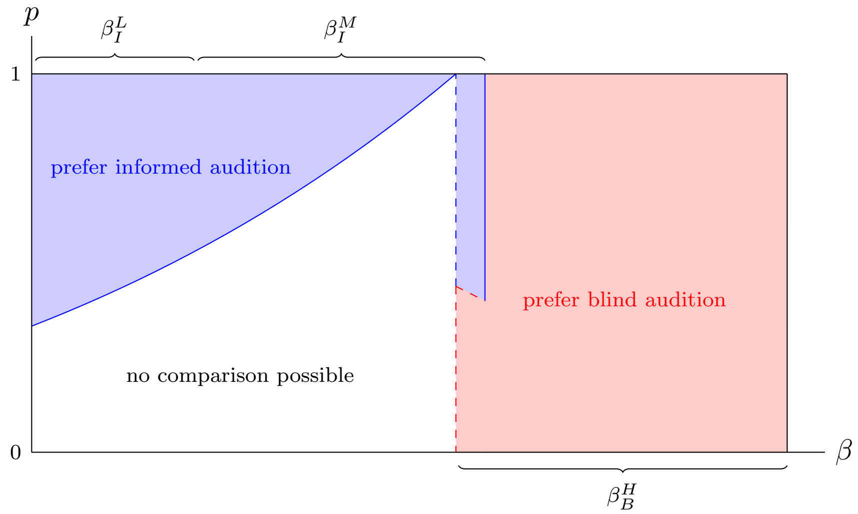


Figure 2.3: Evaluator's preferences over auditions for different biases and priors.

2.5.1 Applicant Preferences over Auditions

To complement the discussion of the evaluator's optimal audition form, I consider the applicant's audition preferences. In particular, having solved for optimal effort levels in both audition forms, I can conclude under which conditions on the bias and prior each type prefers a blind or an informed audition to maximise her expected utility.

Proposition 5. *(i) A low-ability female prefers a blind audition if the evaluator's bias against female applicants is low; that is, $\beta \in \beta_B^L$. (ii) She is indifferent if the evaluator's bias against female applicants is high; that is, $\beta \in \beta_B^H$.*

Intuitively, a low-ability female benefits from the move to a blind audition under the condition that the applicant pool is predominantly of high ability and the evaluator's bias is sufficiently low; for example, she benefits for biases weakly below $\beta(0.5) \approx 0.37$ when the prior is one-half (Figure 2.4a). It allows her to pool with high-ability applicants, concealing her low ability and gender, and, thus, be hired with nonzero probability. In an informed audition, she would never participate because having to reveal both gender and ability would play out against her. When the evaluator's bias is instead high, she is indifferent since her best response is not to participate in either audition form.

Proposition 6. *(i) A high-ability female prefers an informed audition if the evaluator's bias against female applicants is low. (ii) She prefers a blind audition if the evaluator's bias against female applicants is high; that is, $\beta \in \beta_B^H$.*

To gain some insight, fix a prior that is sufficiently high such that the range of biases for which a high-ability female prefers an informed audition is determined by the intersection of

U_B and U_I (Figure 2.4b). Then, as the prior that the applicant is of high ability approaches one, the range of biases for which a high-ability female prefers an informed audition is shrinking and approaches zero. Intuitively, for this applicant, it becomes relatively more important to conceal her gender as the evaluator already knows that she is likely to be of high ability. Submitting a CV prior to the audition is of little value to her.

Proposition 7. *(i) A low-ability male prefers a blind audition if the evaluator's bias against female applicants is low. (ii) He prefers an informed audition if the evaluator's bias against female applicants is high; that is,* $\beta \in \beta_B^H$.

For a very low bias, the intuition behind Proposition 7 is akin to that for low-ability females. A low-ability male benefits from the move to a blind audition under the condition that the applicant pool is predominantly of high ability and the evaluator's bias is sufficiently low (Figure 2.4c). For this applicant, it is attractive to conceal his low ability in a pooling blind audition. He is, in essence, benefiting from an informational externality because the applicant pool is increasingly able. However, beyond a critical value of the bias, he trades off this advantage and rather identifies himself as low-ability male to reap the benefits of the evaluator's bias against female applicants.

Proposition 8. *A high-ability male prefers an informed audition for all levels of bias.*

Intuitively, a high-ability male never wants to conceal either his ability or gender, because his type always plays out in his favour. Under the condition that the evaluator's bias is strictly positive, even if the applicant pool was predominantly of high ability, he would benefit from submitting his CV to identify himself as a male. Graphically, this means that, for any strictly positive bias, U_B will always be below U_I, even if the prior approaches one (Figure 2.4d). This result contrasts with high-ability females whose preferences change to a blind audition as the applicant pool becomes increasingly skilled.

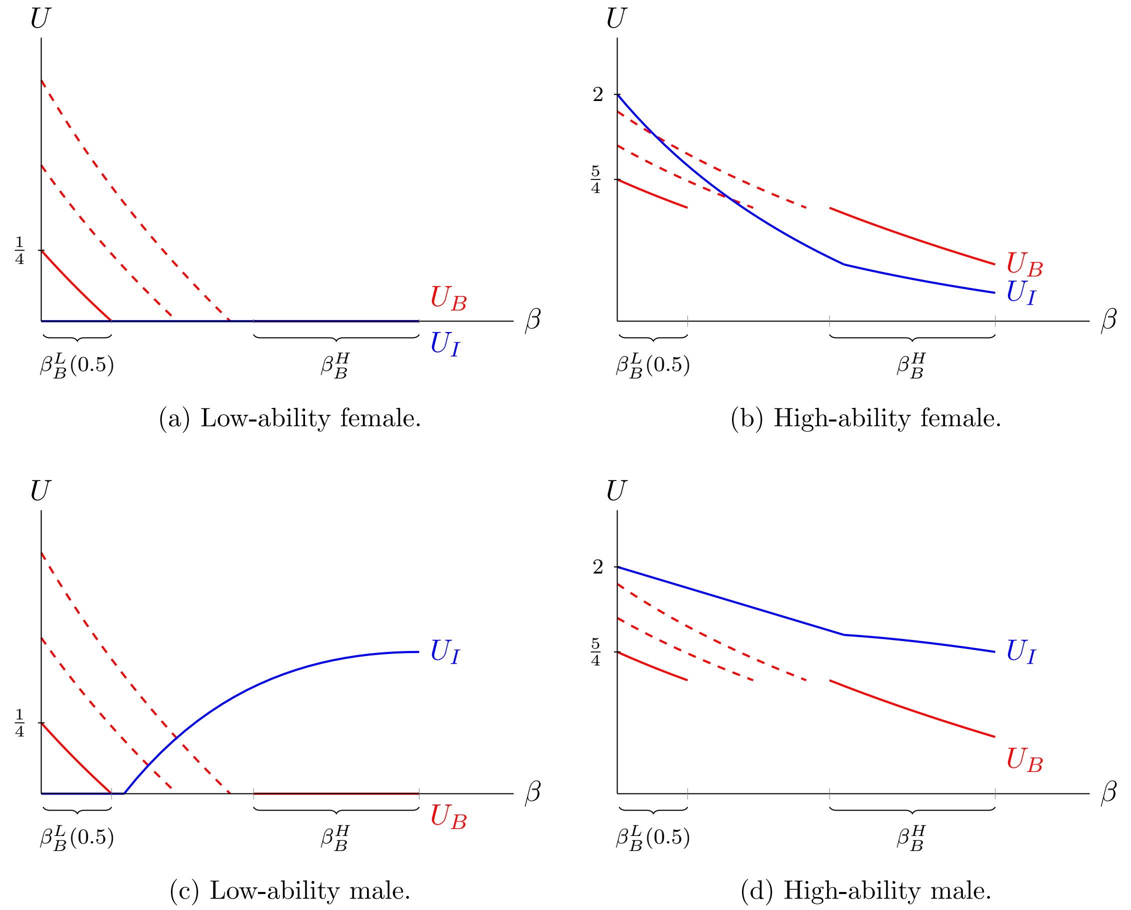


Figure 2.4: Expected utilities in a blind and informed audition.

2.5.2 Discussion of Results

Having discussed the preferences of the four applicant types, I conclude that, if the evaluator's bias is low, applicants' audition preferences differ along the ability dimension of their types. The highly able prefer to identify themselves as such in an informed audition whereas the less able prefer to pool with them in a blind audition. In contrast, if the evaluator is highly biased, applicants' audition preferences differ along the gender dimension. Males prefer an informed audition as the bias plays out in their favour. High-ability females, on the other hand, prefer a blind audition: they rather trade off not being able to reveal their high ability for not having to reveal their gender.

Not surprisingly, the evaluator's preferences align with the preferences of high-ability applicants if his bias against females is low. The evaluator wishes to screen applicants by their ability. In this case, requiring a CV results in targeted effort incentives. High-ability applicants favour this policy as they want to distinguish themselves from low-ability applicants. More surprisingly, the evaluator's preferences align with the preferences of high-ability females if his bias is high. In this case, to provide targeted effort incentives, the evaluator does not require a CV and holds the audition behind a curtain. This deters low-ability applicants and, as a side effect, plays out in favour of high-ability females.

The predictions of my model are qualitatively in line with Goldin and Rouse's (2000) findings that "the switch to blind auditions can explain 30 percent of the increase in the proportion female among new hires and possibly 25 percent of the increase in the percentage female in the orchestras from 1970 to 1996" (p.738). Yet, my model sheds light on the underlying forces. First, if there is at most a low bias against female applicants, such that audition preferences differ along the ability dimension, the authors' findings may be driven by the possibility of hiring low-ability females in a blind audition. In fact, the nonzero probability of hiring low-ability females outweighs the reduced probability of hiring high-ability females in a blind audition when the bias is low. To see this, note that the reduced probability is essentially the driver for why high-ability females prefer an informed audition, as their effort level remains unchanged. In other words, for the same effort cost they are less likely to be hired in a blind audition. Second, if the bias is high, such that audition preferences differ along the gender dimension, their findings may be driven by the increased probability of hiring high-ability females in a blind audition, with no change for their low-ability counterparts.

2.6 Auditions under Performance Uncertainty

Does a highly biased evaluator still prefer relying on a blindfolded performance if there is randomness in the environment? In the performing arts, such randomness is common; for example, due to varying temperature and humidity of the venue (Levinson, 2017). Hence, it is important to understand whether the evaluator's audition preferences are affected by how strongly the applicant's performance quality correlates with her effort. I, thus, relax Assumption 4 in this section and provide an extension to asymmetric performance uncertainty.

Assumption 5. $\Pr(q_H|e_H) = 1$ *and* $\Pr(q_L|e_L) = 1 - \varepsilon$.

Performance quality is stochastically determined by the applicant's effort and one additional uncertainty parameter $\varepsilon \in [0, \frac{1}{2})$. As this parameter gets close to zero, the extended model approaches the benchmark model.

The asymmetry is motivated by the observation that high effort in the performing arts commonly takes the form of deliberate practice (Ericsson et al., 1993; Ericsson, 2006), and learning about performance-controlling factors (McPherson and Schubert, 2004). Performers often describe the process as one of overlearning and overpreparation (Hays and

Brown, 2006). Thus, deliberate practice removes the element of randomness in the mapping from high effort to high performance quality because it allows "the performance to become sufficiently part of oneself such that the response becomes automatic, regardless of what happens" (ibid., p.99). Low effort, in contrast, does not have a deterministic outcome. In this case, "heat, light, noise, as well as any other conditions that might interfere with obtaining a fair assessment of an individual's performance" (Castiglione, 1985, p.34) can play a major role. Assumption 5 may also capture the inherent subjectivity in performance evaluation: the evaluator can identify who engaged in deliberate practice but his standards are too lenient to always identify at the margin those applicants who did not.[23]

2.6.1 Comparison of Auditions under Performance Uncertainty

I solve the model under performance uncertainty and discuss the effect on the evaluator's net utility in both audition forms in Section A.1.2. In what follows, I summarise under which conditions on the bias, the prior and the degree of performance uncertainty my results on the evaluator's preferences continue to hold.

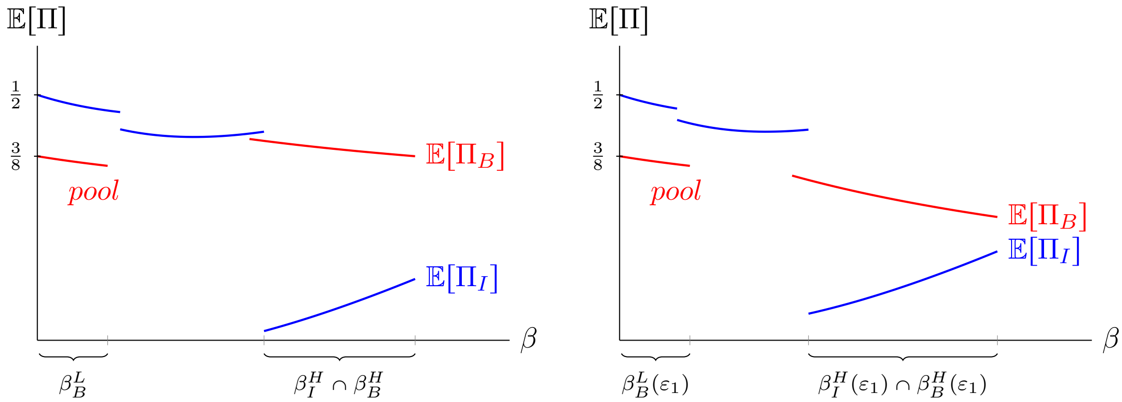
First, given performance uncertainty above a certain threshold, full separation in a blind audition is not possible. There is *partial separation*: highly able applicants exert high effort and the less able exert low effort. Therefore, this audition form does not have the same fully fledged effect as under certainty; that is, a blind audition cannot provide targeted effort incentives when the evaluator is highly biased. This dampens the evaluator's expected net utility. For a fixed prior, compare, for example, the evaluator's full separation net utility when $\varepsilon_0 = 0$ in Figure 2.5a with his partial separation net utility when $\varepsilon_1 = \frac{1}{10}$ in Figure 2.5b: while a highly biased evaluator still finds a blind audition more profitable for this ability composition of the applicant pool and this degree of performance uncertainty, the expected net gain is smaller.

Second, with uncertainty exceeding the threshold, blind auditions are not guaranteed to be more profitable than informed auditions when the evaluator is highly biased (Figure 2.6). However, blind auditions are still guaranteed to be more profitable for a sufficiently high prior approximately greater than 0.46. Intuitively, while the less able participate in the blind audition, the highly biased evaluator can, at least, provide effort incentives for close to the majority of the highly able.

Third, full separation insured the highly biased evaluator against negative net utility in a blind audition. With uncertainty above the threshold and partial separation, the

[23] A different interpretation is that the applicant may be lucky, or have mastered the "inner game" (Gallwey, 1974).

evaluator's net utility may be negative for very low priors. In fact, there are bias-prior-uncertainty combinations for which neither blind nor informed auditions are profitable. For $\varepsilon_1 = \frac{1}{10}$, these combinations are the grey-shaded areas in Figure 2.6.



(a) $\Pr(\eta_H) = \frac{1}{2}$ and $\varepsilon_0 = 0$.

(b) $\Pr(\eta_H) = \frac{1}{2}$ and $\varepsilon_1 = \frac{1}{10}$.

Figure 2.5: Evaluator's expected net utility in a blind and informed audition under performance uncertainty.

To gain some intuition behind the evaluator's distribution of preferences, suppose that, if the degree of performance uncertainty ε_1 is moderate, the evaluator's bias and the prior on high ability are drawn uniformly from their respective supports. Thus, all points in Figure 2.6 are equally likely before the three-stage game under performance uncertainty is played. As in Section 2.5, I discount the white areas for which no comparison is possible when calculating the ex-ante percentage with which a blind audition (red area) is preferred relative to an informed audition (blue area), or neither (grey area). The evaluator prefers a blind audition approximately fifty percent and an informed audition approximately forty-five percent of the time. Neither audition is preferred approximately five percent of the time (Figure 2.6). First, this highlights the sizeable risk of market failure under performance uncertainty: the evaluator might prefer to not hold an audition at all. Second, this highlights that an informed audition gains only about nine while a blind audition loses about fourteen percentage points in the move from a certain to a moderately uncertain environment. In other words, the evaluator is worse off as the performance uncertainty reduces the attractiveness of a blind audition by more than it increases the attractiveness of an informed audition. Taken together, the welfare loss of asymmetric performance uncertainty comes from the possibility of market failure: the no-audition scenario did not exist under Assumption 4 in Figure 2.3. Further, the combination of moderate performance uncertainty with a simple form of heterogeneity in the bias across evaluators as well as in the ability composition of applicant pools may rationalise the co-existence of blind and informed auditions in more equal proportions.

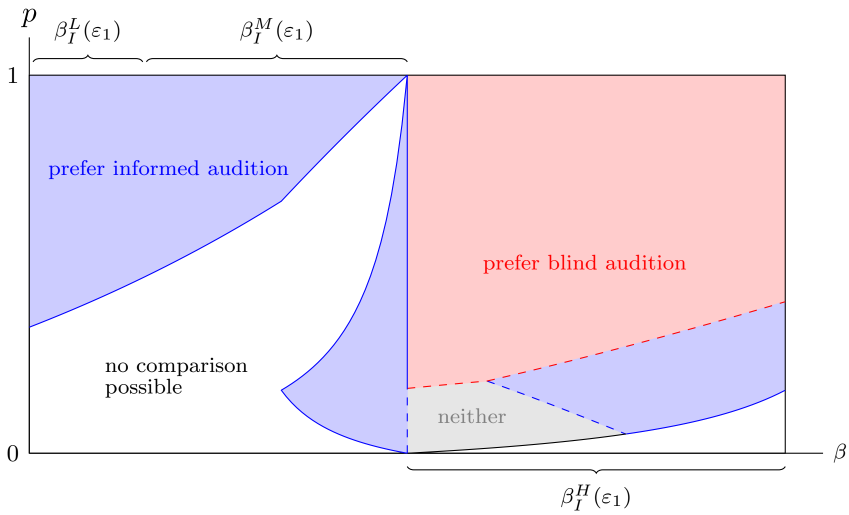


Figure 2.6: Evaluator's preferences over auditions for different biases and priors for $\varepsilon_1 = \frac{1}{10}$.

2.6.2 Policy Recommendations under Performance Uncertainty

If performance uncertainty is an inherent feature of the environment and cannot be alleviated, then policy should ensure that the applicant pool is of sufficiently high ability. Examples could be early childhood interventions to identify talent and to build performance excellence gradually via deliberate practice and, thereafter, merit scholarships for formal training to secure a steady influx of highly able musicians into the applicant pool. Direct interventions could target artistic poverty and the necessity for multi-jobbing, often in teaching, to earn a subsistence income as main barriers to human capital investment later in life. One could ask whether "auditioning, networking, keeping fit, practising and unpaid rehearsing [should] be counted as part of the artists' labo[u]r supply" (Towse, 1996, p.102) to reward human capital investment.

Under performance uncertainty, such ability-targeting policy is beneficial for two reasons: first, it guarantees the profitability of the partially separating blind audition when the evaluator is highly biased. Second, it guarantees that blind auditions outperform informed auditions. This may be desirable for equality of opportunity. Intuitively, under performance uncertainty, an applicant pool of sufficiently high ability helps counteract that blind auditions only achieve partial separation. Furthermore, it helps avoiding market failure, which may arise if neither informed nor blind auditions are profitable. In this case, it would seem natural for the evaluator to revert to the outside option immediately without holding an audition at all. This would lead to continual subcontracting or zero-hour contracts in the performing arts, reinforcing gender inequality via the difficulty of combining precarious employment with motherhood and caring responsibilities (Conor et al., 2015).

The recommendation for government to stimulate the acquisition of skills is in line with existing research on effective interventions in the presence of low-skill bad-job equilibria (Snower, 1994): when there is a small proportion of highly able types, the evaluator has little incentive to fill the position permanently, which, in turn, reinforces the underprovision of skilled labour. Furthermore, a training supply externality arises because an increase in the number of high-ability types in the population raises the probability that the evaluator will face and hire a highly able applicant, which raises his expected payoff from holding the audition. There may also be a rationale for speed if market failure leads the intervention to be less effective (Oberholzer-Gee, 2008). The policy recommendation to promote efficiency and equity under performance uncertainty finds long-standing empirical support in a number of countries: state-run music conservatoires and departments of music and drama as part of more general educational institutions provide education opportunities. Training orchestras support the professional aspects of the performing arts (Throsby and Withers, 1979).

Market failure may have adverse and wide-ranging consequences for welfare beyond gender inequality. One mechanism could be reduced civic engagement. Research by the National Endowment For The Arts (2009) finds that"Americans who create or perform art are more civically active than the general U.S. adult population" (p.6). A second mechanism could be reduced spillover effects accruing to the neighbourhood in which musicians reside and work. Research by the Social Impact of the Arts Project (2017) at the University of Pennsylvania, School of Social Policy and Practice, finds that cultural resources positively affect a neighbourhood's health and crime rate, and the outcomes of its schools. While this is not a causal relationship, cultural resources, such as artists, are argued to be vital for an environment that promotes social wellbeing.

2.7 Blind Audition vs Affirmative Action

A blind audition serves as a policy to counteract first impression biases. As a measure that eliminates the effect of the evaluator's taste for male applicants on his hiring decision, it creates an environment of *equal opportunity* for females and males: conditional on ability and performance quality, females and males have the same hiring probability. Yet, government policy often goes further, imposing *affirmative action*: "any measure, beyond simple termination of a discriminatory practice, adopted to correct or compensate for past or present discrimination or to prevent discrimination from recurring in the future" (U.S. Commission on Civil Rights, 1977, p.2). Examples are reverse discrimination, preferential treatment and quotas. A natural question is, therefore, how blind auditions as an equal opportunity policy compare to these more drastic policies aimed at mitigating biases.

2.7.1 Simple Quota

A number of governments impose quotas to increase female representation in the labour market. Since 2006, for example, Norway enforces a previously voluntary quota of 40 percent for the boards of directors of publicly listed companies (Bertrand et al., 2019). In my benchmark model, such a simple quota may be implemented as the constraint that a female applicant is guaranteed to be hired with some probability $\underline{x} \leqslant \frac{1}{2}$. This puts a floor on the hiring probabilities of females in an informed audition:

$$\begin{aligned}
\Pr(h|q_L, \eta_L, f) &= \underline{x}, \\
\Pr(h|q_H, \eta_L, f) &= \max\left\{\underline{x}, \frac{1}{2+\beta}\right\}, \\
\Pr(h|q_L, \eta_H, f) &= \max\left\{\underline{x}, \frac{1}{2+\beta}\right\}, \\
\Pr(h|q_H, \eta_H, f) &= \max\left\{\underline{x}, \frac{2}{2+\beta}\right\} = \frac{2}{2+\beta} \quad \forall \quad \beta \in [0,2).
\end{aligned}$$

Specifically, if $\underline{x} = \frac{4}{10}$, then low-ability females exerting low effort are always hired with this probability; that is, the quota always binds. Furthermore,

$$\Pr(h|q_H, \eta_L, f) = \Pr(h|q_L, \eta_H, f) = \begin{cases} \frac{1}{2+\beta} & \text{if } \beta \leqslant \frac{1}{2}, \\ \frac{4}{10} & \text{otherwise.} \end{cases}$$

As a result, the simple quota affects the participation and incentive compatibility constraints of female applicants. In particular, for low-ability females,

$$U(e_L, \eta_L, f) = \frac{4}{10}w - \frac{1}{2}, \qquad U(e_H, \eta_L, f) = \begin{cases} \frac{1}{2+\beta}w - 2 & \text{if } \beta \leqslant \frac{1}{2}, \\ \frac{4}{10}w - 2 & \text{otherwise,} \end{cases}$$

where $w = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. The quota alters the participation constraint of low-ability females: low effort yields a positive payoff for all biases. High effort, in contrast, always yields a negative payoff. A policy of guaranteeing females a hiring probability of 40 percent, therefore, induces low-ability females to exert low effort for all levels of bias, and forces a loss on the evaluator in this subgame. What is more, a simple quota may go beyond equal opportunity for low-ability females and males; that is, for $\beta < \frac{4}{3}$, it results in reverse discrimination or preferential treatment:

$$\Pr(h|q_L, \eta_L, m) = \frac{\beta}{2+\beta} < \frac{4}{10} = \underline{x} = \Pr(h|q_L, \eta_L, f) \text{ if } \beta < \frac{4}{3}.$$

For high-ability females,

$$U(e_L, \eta_H, f) = \begin{cases} \frac{1}{2+\beta}w - \frac{1}{4} & \text{if } \beta \leqslant \frac{1}{2}, \\ \frac{4}{10}w - \frac{1}{4} & \text{otherwise}, \end{cases} \qquad U(e_H, \eta_H, f) = \frac{2}{2+\beta}w - 1,$$

where $w = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. The quota alters the incentive compatibility constraint of high-ability females: low effort yields a higher payoff than high effort for biases exceeding $\frac{51-\sqrt{1929}}{8} \approx 0.89$. As a result, the threshold bias beyond which effort incentives for a highly able female are distorted is lower. For moderate biases, $\beta \in (\frac{51-\sqrt{1929}}{8}, \frac{6}{5}]$, the quota, thus, has unintended consequences: highly able females are hired less often due to the quota's effect on effort incentives. Moreover, this hiring probability is precisely the quota for such moderate biases (Figure 2.7).

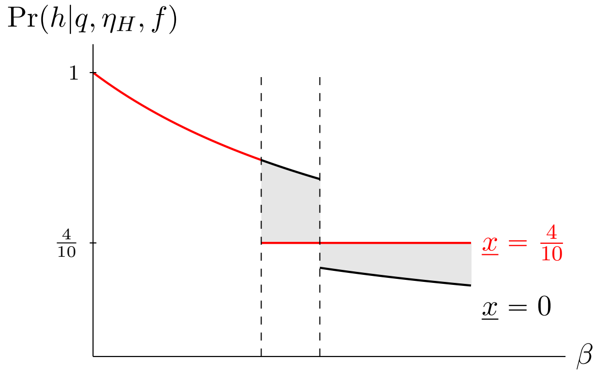


Figure 2.7: Effect of a simple quota on the hiring probability of a high-ability female via her effort response.

In consequence, simple quotas cannot provide equal opportunity and targeted effort incentives in my benchmark model. Moreover, my results echo the findings of Coate and Loury (1993) and Moro and Norman (2003) in that quotas can hurt the intended beneficiaries. In my setting, the quota causes the hiring probability of highly able females to fall for a moderate range of biases before it achieves its aim of putting a floor on the hiring probability for higher biases.

2.7.2 Beyond Simple Quotas

Simple quotas are not sufficient for equal opportunity and targeted effort incentives when they affect the behaviour of the target group and, as a consequence, the behaviour of the evaluator. An alternative is to force the evaluator to hire females a priori (across subgames) with some probability $\underline{x}$. The evaluator then has discretion to adjust the hiring probabilities within subgame (η_L, f) and (η_H, f) to provide targeted effort incentives. In particular, to avoid attracting low-ability females, he can hire exclusively highly able females more often than demanded by the hiring rule (2.3) in order to fulfil the quota. Given effort responses in stage 2, the probability of hiring female applicants across subgames is

$$\Pr(h|f) = \begin{cases} \frac{2p}{2+\beta} & \text{if } \beta \leqslant \frac{6}{5}, \\ \frac{p}{2+\beta} & \text{otherwise.} \end{cases}$$

For low biases, therefore, the quota is binding if $\underline{x} > \frac{2p}{2+\beta}$. For example, when $\underline{x} = \frac{4}{10}$ and $p = \frac{1}{2}$, the quota is binding for biases exceeding one-half, and the evaluator is forced to hire highly able females with probability $\underline{x}/p$ in subgame (η_H, f) to fulfil the quota. As a corollary, such a quota is only feasible if it does not exceed the prior that the applicant is of high ability.

2.7.3 Employment Subsidy

An alternative practice is to pay subsidies to firms who employ a certain type of worker (e.g., Card et al., 2010, 2017). In my benchmark model, this may be implemented as a subsidy of at most β conditional on hiring a female applicant in the audition. The outside option remains unchanged; that is, the evaluator does not receive the employment subsidy if he hires a substitute that turns out to be a female. The evaluator's gross utility (2.1) changes to

$$V(q, \eta, g) + \alpha\beta g = f(q, \eta) - (1 - \alpha)\beta g,$$

where $\alpha = 1$ represents a full employment subsidy. Given the altered gross utility and no change in the outside option, as $\alpha \uparrow 1$, the conditional hiring probabilities of females approach the ones of males. In fact, under a full subsidy, if $\beta \in \beta_I^L$, only high-ability applicants participate and exert high effort. If $\beta \in \beta_I^M$, low-ability applicants also participate and exert low effort. If $\beta \in \beta_I^H$, all applicants participate and exert low effort. For low biases, the full employment subsidy can, therefore, achieve both targeted effort incentives and equal opportunity. For moderate and high biases, equal opportunity is achieved at the expense of attracting low-ability females and males alike.

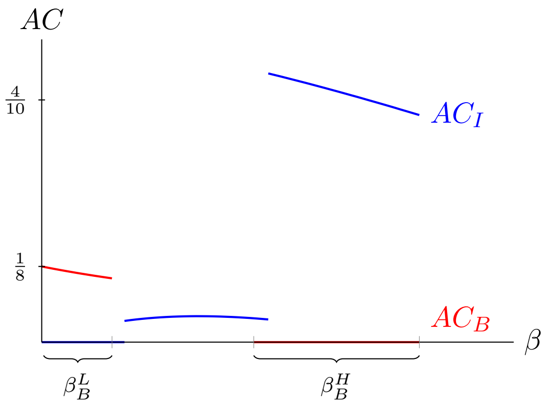
2.8 Agency Cost

Equal opportunity and efficiency are commonly argued to be conflicting policy objectives (e.g., Sowell, 2004). To analyse whether this trade-off arises under taste-based discrimination, I measure efficiency through the evaluator's agency cost in the two audition forms.[24] In particular, for a given bias, this agency cost is the difference between the evaluator's

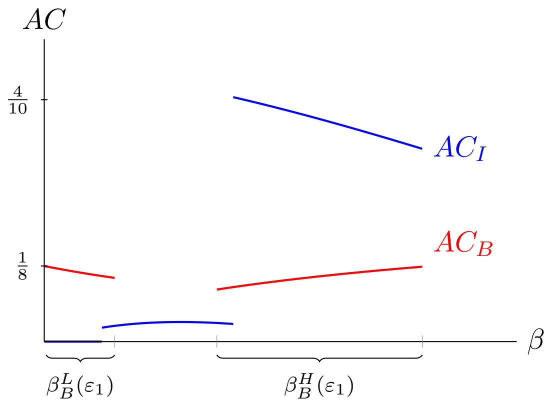
[24]In the literature, agency cost is defined as the sum of the principal's monitoring expenditures, the bonding expenditures by the agent, and the residual loss that arises from a divergence between the agent's decisions and those decisions which would maximise the principal's welfare (Jensen and Meckling, 1976). In my model, the agency cost is purely driven by the residual loss arising from distorted effort incentives.

hypothetical payoff when only highly able applicants participate and exert high effort in the audition form and the evaluator's actual expected net utility. In an informed audition, the hypothetical payoff is derived from extrapolating $\mathbb{E}[\Pi_I|\beta \in \beta_I^L]$ to higher biases. In a blind audition, the hypothetical payoff is derived from extrapolating $\mathbb{E}[\Pi_B|\beta \in \beta_B^H]$ to lower biases in the absence of performance uncertainty.[25]

In the absence of performance uncertainty, if the evaluator's bias against female applicants is low, the agency cost is zero in an informed audition due to targeted effort incentives. In contrast, when the bias is moderate, there is an agency cost from attracting low-ability males who exert low effort. When highly biased, the agency cost increases further due to the provision of low rather than high effort incentives for highly able applicants. In a blind audition, the agency cost is zero if the evaluator's bias is high. Conversely, when the bias is low, an agency cost arises from attracting also low-ability applicants who exert high effort. Figure 2.8a depicts the agency cost (AC) in both audition forms for a prior of one-half.



(a) $\Pr(\eta_H) = 0.5$ and $\varepsilon_0 = 0$.



(b) $\Pr(\eta_H) = 0.5$ and $\varepsilon_1 = \frac{1}{10}$.

Figure 2.8: Agency cost in a blind and informed audition.

Consequently, above a certain threshold bias, a blind audition is an efficient mechanism to screen applicants by ability and to provide targeted effort incentives. What is more, equal opportunity and targeted effort incentives are complementary objectives: there is no trade-off between implementing blind auditions as a policy to counteract first impression biases and efficiency due to adverse effects on targeted effort provision. For a low bias, in contrast, an informed audition poses an efficient mechanism. In this case, equal opportunity and targeted effort incentives are conflicting objectives: while blind auditions serve as a policy to counteract biases, they may lead to an avalanche of applicants and efficiency loss as unintended consequences.

[25] A formal definition of agency cost and a formal statement of this section's results is in Section A.1.3.

Under asymmetric performance uncertainty, however, a blind audition no longer poses an efficient mechanism: partial separation leads to a positive agency cost for a highly biased evaluator (Figure 2.8b). Given that this agency cost is decreasing in the prior, ability-targeting interventions reduce the inefficiency. The agency cost in an informed audition for a moderate and high bias is lower relative to certainty. Intuitively, the evaluator prefers to hire an applicant who delivers a high-quality performance. Under certainty, this objective coincides with the provision of high effort incentives. Under performance uncertainty, this is no longer the case and the agency cost from a low effort decision is reduced due to the possibility that low effort still results in a high-quality performance. Due to the asymmetry in Assumption 5, the agency cost in the high-effort pooling equilibrium remains unchanged.

2.9 Gender-blind Curriculum Vitae

In my model, concealing the applicant's gender necessitates concealing the applicant's ability. Even with such a trade-off, I show that a blind audition can be optimal for the evaluator. Yet, with advanced technology and software able to edit out selected pieces of information on a CV, the evaluator could design the blind audition: he may choose to neither learn ability nor gender, or ask for a gender-blind CV that contains valuable information about ability.

Therefore, an important question for policy is whether a gender-blind CV coupled with a blindfolded performance can serve as an improved equal opportunity policy. In other words, if the evaluator was bound to hold a blind audition to create an environment of equal opportunity, would he prefer to see a gender-blind CV prior to the blindfolded performance?[26] If the evaluator decides to obtain information about ability beforehand, he preempts signalling: given a gender-blind CV, low-ability applicants cannot pool with the highly able in the blind audition.[27] The submission of a gender-blind CV prior to the blindfolded performance motivates highly able applicants to exert high effort for biases below a certain threshold. Further, low-ability applicants never participate when the evaluator requires a gender-blind CV. Consequently, for low biases, a blind audition coupled with a gender-blind CV does not have the unintended consequences of a blind audition that conceals both gender and ability.

The addition of gender-blind CVs to a blind audition reverses the results in Lemma 3: a blind audition without any form of CV provides targeted effort incentives for high biases; a blind audition with gender-blind CVs provides targeted effort incentives for low biases.

[26] A formal statement of this section's results is in Section A.1.4.

[27] As before, gender cannot be signalled so that the evaluator will always believe the applicant to be a female or a male with equal probability.

Thus, if the evaluator was bound to hold a blind audition for equal opportunity, he would choose to see a gender-blind CV for a low bias, $\beta \in \beta_B^L|_{p\uparrow 1}$. He would be indifferent for a moderate bias, $\beta \in \beta_G^L \cap \beta_B^H$. For a high bias, $\beta \in \beta_G^H$, he would choose not to see a CV in any form. A highly biased evaluator takes advantage of the need to signal ability in a blind audition without any form of CV: the belief that a low-quality performance comes from a less able applicant motivates the highly able to exert high effort.

Given an optimally designed blind audition, I can compare the evaluator's expected net utility across the blind and informed audition for essentially all bias-prior combinations (Figure 2.9a). If the evaluator's bias against female applicants is low, he prefers an informed audition over a blind audition with ability revealed. For high biases, the evaluator prefers a blind audition with ability concealed over an informed audition. For moderate biases, the prior about ability determines whether the evaluator prefers an informed or a blind audition with ability revealed through gender-blind CVs. Consequently, Proposition 4 is robust to the introduction of audition design in stage 1. Further, the evaluator's distribution of preferences is similar if he can design the blind audition: the optimally designed blind audition is preferred approximately sixty-two percent of the time. These insights extend to performance uncertainty when a blind audition with ability concealed cannot provide targeted effort incentives (Figure 2.9b).

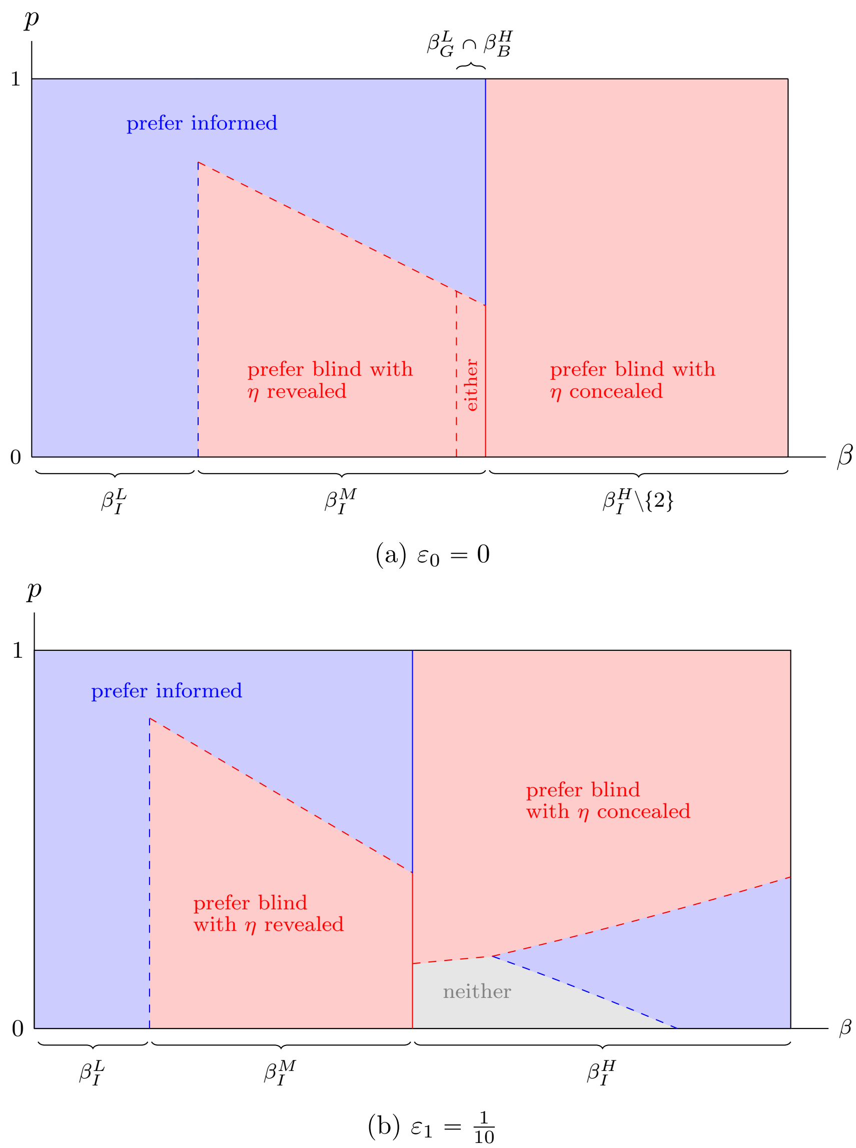


Figure 2.9: The effect of gender-blind CVs on the evaluator's preferences over auditions.

2.10 Conclusion

In this paper, I developed a game-theoretic framework in which the evaluator is biased against females and his bias is common knowledge. I showed that, above a threshold bias, the evaluator prefers a blind audition to provide high effort incentives exclusively for high-ability applicants. I also showed that a highly biased evaluator's preferences align with those of a highly able female. The introduction of performance uncertainty may lead to market failure or may render informed auditions more profitable, rationalising ability-targeting interventions. While a full employment subsidy can be a viable alternative to a blind audition for low biases, a simple quota cannot provide equal opportunity and targeted effort incentives in my benchmark model. Finally, I showed that it can be optimal to reveal ability through a gender-blind CV.

My paper opens up intriguing avenues to study in more depth the potential trade-off between policies mitigating biases and effort incentives for applicants. In Section 2.7, I have taken a first look at quotas and subsidies. Another example is a setting in which the evaluator's bias is not common knowledge: what if applicants are uncertain or underestimate the evaluator's bias? Specifically, applicants perceiving their identity more strongly may attend to the evaluator's bias more (Antecol and Cobb-Clark, 2008). What if the evaluator is partially sophisticated or naive about his bias when choosing the audition form? This enquiry is motivated by evidence that many forms of bias work at a subconscious level; for example, via a vision heuristic (Footnote 8). Few studies examine the implications of competition among applicants, allowing for potentially more complex biases in an informed audition (Page and Page, 2010). A first point of departure may be a tournament model à la Cornell and Welch (1996). Another promising avenue is to microfound the evaluator's bias; for example, via a categorical model of cognition (Fryer and Jackson, 2008). Finally, future research could consider situations in which deliberate practice does not allow an applicant to deliver a high-quality performance with certainty; overlearning and overpreparation could arguably make an applicant more adept at responding to factors outside of her control.

Chapter 3

The Handmade Effect: A Model of Conscious Shopping in an Industrialised Economy*

3.1 Introduction

> *"We believe that human connection is central to buyer engagement. On Etsy, we emphasize that the items listed for sale are brought to life by real people."*
>
> – Etsy, Inc., 2018 Annual Report.

The e-commerce website Etsy was founded in 2005 as a marketplace for handmade products. In an era of technological advancement, Etsy's business model is notable: Everything that is listed in the marketplace must be "handmade or unique and assembled with production partners, vintage or craft supplies" (Etsy, Inc., 2019a, p.11). In a wide range of categories – including jewelry, clothing, toys and furniture – the sellers' production processes could readily be automated, which could potentially result in improved quality and fewer defective products. Nevertheless, the number of buyers and sellers on Etsy has been rising with gross merchandise sales (GMS) surpassing 2.5 billion U.S. dollars in the second quarter of 2020 (Figure 3.1). It is, therefore, worth understanding why this sector of the economy can persist in the twenty-first century.

*Published (excl. additional materials in Section A.2) in the *Review of Industrial Organization* on 2 November 2021.

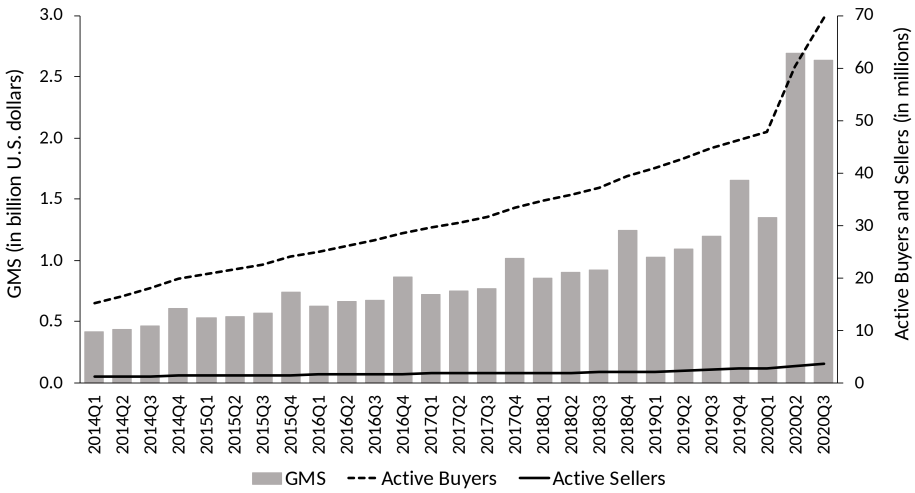


Figure 3.1: Business activity on Etsy since 2014 (data source: Etsy, Inc.).

In this paper, I offer an explanation for the observed puzzle. I argue that a handmade effect on the consumer side (Fuchs et al., 2015) can explain why some firms – like those selling on Etsy – do not adopt superior production processes and instead choose to produce and advertise their products as handmade. I present a model that features a unit mass of shoppers. They come in two types: quality shoppers, and conscious shoppers. In this setting, product differentiation has two dimensions (e.g., Economides, 1989): All shoppers value product quality equally, but they differ in the weight that they give to the production process. From a firm's perspective, there are nontrivial interactions: Hand production comes at a higher effort cost, and may be inferior at translating any given effort into quality.

I show that when a fraction of shoppers is sufficiently concerned about the conditions under which the product is manufactured and, therefore, is willing to pay a handmade premium, the firm adopts a handmade strategy. A minimum quality standard, however, can alter the firm's behaviour and induce switching to machine production. When hand production is associated with increasing marginal cost, I show that a niche handmade product can be perceived to be of higher quality than a mass market machine-made product.

I extend the model to settings in which two ex-ante identical firms decide simultaneously on the production process, pricing strategy, and per-unit effort. I show that the existence of sufficiently conscious shoppers can explain the specialisation of firms and, thus, the observed co-existence of handmade and machine-made products in the economy. I show that such specialisation is efficient, and robust to collusion if the proportion of conscious shoppers in the market is not too large and their concern about the conditions under which the product is manufactured is not too high.

I then introduce conformist shoppers who behave exactly like conscious or quality shoppers depending on their beliefs. In this setting, a monopolist may be able to use a handmade strategy to signal a social norm of conscious consumption. When credible, this strategy generates a crowding-in effect and explains sales growth even in markets where conscious shoppers are believed to be in the minority. Finally, I show that my model can shape intuitions also on ethical production, and can explain the mainstreaming of Fair Trade in Europe in the late 1990s (Renard, 2003) through such a crowding-in effect.

3.1.1 Related Literature

My paper relates to the growing economic literature on ethical consumption (e.g., Stiefenhofer, 2019). The empirical literature has focused on estimating consumers' willingness to pay a premium for ethical products (e.g., Galarraga and Markandya, 2004; Loureiro and Lotade, 2005; Arnot et al., 2006). For example, when buying more expensive Fair-Trade products, consumers reveal their preferences for ethical products. They are inferred to obtain additional utility from the product's Fair-Trade label. My model builds on these insights: conscious shoppers gain utility from buying a handmade product whereas quality shoppers care only about product quality.

My extension to two firms relates to how product differentiation can relax price competition. In his classical work on monopolistic competition, Chamberlin (1933) notes that product differentiation leads buyers to pair with sellers according to their preferences. Firms evade price competition by focusing buyers' attention towards a trade-mark, or by competing on the basis of quality. In my model, attention is focused on the production process: A handmade strategy relaxes price competition. Champsaur and Rochet (1989) obtain a similar result by allowing firms to choose non-overlapping qualities followed by prices. Sebald and Vikander (2019) consider a related signaling game in which a monopolist faces a fraction of consumers who care about what other consumers believe.

The rest of the paper is organised as follows: Section 3.2 introduces the model. I discuss my main results when the firm acts as a monopolist in Section 3.3, and when the firm is part of a duopoly in Section 3.4. In Section 3.5, I explain how a monopolist may use a handmade strategy to signal a social norm of conscious consumption. Section 3.6 concludes with an application to Fair Trade, and avenues for future research.

3.2 Model

The key mechanism in my model is the *handmade effect*: Some consumers perceive that handmade products symbolically contain the producer's love for the product and its production process (Fuchs et al., 2015). In particular, they may value that the product is crafted and touched by someone with a passion for the product (Argo et al., 2008). Moreover, the handmade label imbues the product with the producer's personality or "face" (Johnston and Baumann, 2007). The product may be perceived to embody the creator's unique personal expression (Kreuzbauer et al., 2015). Conversely, machine-made products cannot absorb the producer's passion given that the product is often not touched at all during its production (Markoff, 2012).

A second value channel is the "authenticity" of the producer: Some consumers may value that the producer of a handmade product assumes responsibility for the production process; business decisions under machine production do not convey a producer's values but are perceived as marketing strategy (Carroll and Wheaton, 2009; Lehman et al., 2019). I formalise the handmade effect on the consumer side through a fraction C of shoppers being conscious of this handmade effect. These conscious shoppers care about the conditions under which the product is manufactured: They value production by hand h over machine production m. Thus, conscious shoppers base their buying decision on an enlarged set of product attributes compared to the other shoppers (Lancaster, 1966): price, product quality, *and* the production process.

3.2.1 Timing and Actions

The firm F and a unit mass of shoppers (the market) play a two-stage game: In the first stage, the firm chooses the production process $k \in \{h, m\}$, the product's per-unit price p_k, and the per-unit effort level $x_k \geqslant 0$. I assume that product quality is jointly determined by the production process and effort, which I will describe in more detail below.

In the second stage, having observed the firm's decision, the shoppers choose whether to buy one unit of the product. They have an outside option that yields a zero payoff. Let C and $Q = 1 - C$ be the fraction of conscious shoppers and quality shoppers in the market, respectively. Let $a_i \in \{0, 1\}$, $i = C, Q$, be their buying decision.

3.2.2 Firm's Payoffs

For a given k, the firm chooses the per-unit price and the per-unit effort level to solve the profit maximisation problem

$$\max_{(p_k, x_k) \in \mathbb{R}^2_+} \Pi(p_k, x_k; k) = \max_{(p_k, x_k) \in \mathbb{R}^2_+} (a_C C + a_Q Q)\left[p_k - \frac{c_k}{2} x_k^2\right].$$

The firm's per-unit cost is $c(x_k) = \frac{c_k}{2} x_k^2$. Therefore, the firm's profit depends on the cost parameter $c_k > 0$, where $c_h > c_m$. Intuitively, it is more costly to produce one unit of the product by hand. For example, for a given effort level, it requires more time to knit a pair of socks by hand. The assumption that $c_h > c_m$ then captures the firm's opportunity cost of time in a reduced form. Note that for tractability the per-unit cost increases in effort but remains constant in the level of output produced. I will relax this assumption in Section 3.3.1.

3.2.3 Quality Shoppers' Payoffs

Quality shoppers choose to buy one unit of the product if it maximises their utility:

$$\max_{a_Q \in \{0,1\}} V_Q(a_Q; k) = \max_{a_Q \in \{0,1\}} a_Q[q(x_k, k) - p_k].$$

Not buying yields a zero payoff. Conversely, buying one unit yields a payoff of $q(x_k, k) - p_k$. The value $q(\cdot)$ represents the benefit from consuming one unit. Following the literature on vertical differentiation (e.g., Mussa and Rosen, 1978), the benefit rises linearly with the product's quality.[1]

Assumption 1 (Product Quality). $q(x_k, k) = A_k x_k$ *with* $A_m \geqslant A_h \geqslant 1$.

I assume that the product's per-unit quality is increasing at a constant rate A_k in the firm's effort level. The production process, therefore, scales the shoppers' benefit for a given amount of effort, where $q(0, m) = q(0, h)$ and $q(x, m) \geqslant q(x, h)$ for all $x > 0$: Production by hand results in weakly lower product quality and, thus, in a weakly lower benefit for quality shoppers.[2]

This could be due to physical and mechanical variance, or the possibility of human error in the production of h. Compared to the potentially inconsistent quality of handmade products, machine production allows for uniform product quality (Liebl and Roy, 2003), and increased precision (Markoff, 2012). Watches are one example: Machine-made quartz watches are "significantly more precise than the most expensive Swiss mechanical watches made by the hands of a skilled watchmaker in Geneva" (Kreuzbauer et al., 2015, p.764).

Intuitively, the quality shoppers choose to buy one unit of product k if the quality is weakly greater than the price set by the firm in the first stage. What is more, quality shoppers only care about the end product: They shop for the highest quality. Thus, for a given price, they weakly prefer the firm to offer a machine-made product.

Arguably, some handmade products are of higher quality than mass-produced products. One example is the high-value niche wine production in France: Compared to the machine harvesting in the Australian wine industry, France's handcrafted artisan methods yield a

[1] A diminishing marginal benefit, $q(x_k; k)^z$ with $z \in (0, 1)$, will not change the result in Section 3.3 that the firm would sell m to the whole market and h only ever to C as long as z is common to all shoppers. However, it does not always make a handmade strategy relatively more attractive for the firm: A lower z does not necessarily reduce the difference between the optimised profit (or quality) of m and h.

[2] Given $c_h > c_m$, my results do not hinge on $k = h$ being worse at translating any given effort into quality.

limited quantity of high quality (Aylward and Carey, 2009). A second example is Aston Martin: Its limited number of expensive hand-built cars are recognised as authentic and well designed (Carney, 2016). The extension to convex production cost in Section 3.3.1 will accommodate these settings.

3.2.4 Conscious Shoppers' Payoffs

Conscious shoppers choose to buy one unit of the product if it maximises their utility:

$$\max_{a_C \in \{0,1\}} V_C(a_C; k) = \max_{a_C \in \{0,1\}} a_C[q(x_k, k) - p_k + \theta\nu(k)].$$

The exogenous parameter $\theta > 0$ measures their sensitivity to the production process.[3] The designation of a product as handmade to trigger θ for segment C is exogenously determined; for example, through the enforcement of a Handmade Policy (Etsy, Inc., 2019b).

Assumption 2 (One-sided Concern)**.** $\nu(k) = \mathbb{I}(k = h)$.

Let $\mathbb{I}(x = y)$ be an indicator function that is equal to unity if its argument holds true, and zero otherwise. I normalise the conscious shoppers' utility gain from consuming a handmade product to unity. Moreover, I assume that their utility loss from consuming a machine-made product is zero. A one-sided concern of conscious shoppers is a natural assumption given that most firms that use m do not mention the production process explicitly when marketing their product (Fuchs et al., 2015). Consequently, only production by hand is salient when conscious shoppers make the buying decision.

Intuitively, conscious shoppers care about more than product quality: They are also concerned about the conditions under which the product is manufactured. Depending on their sensitivity to the production process, buying one unit of a product that is marketed as handmade may be more important than the product's quality. Carfagna et al. (2014, p.165) provide empirical support for my choice of utility function: Consumers who are strong supporters of the Do-It-Yourself ethic are "less concerned about quality and outcome and more oriented to the process".

3.2.5 Solution Method and Preliminary Results

I solve the model by first considering the shoppers' decision problem. The shoppers buy one unit of the product if $V_i(1; k) \geqslant V_i(0; k)$, $i = C, Q$. Therefore, conscious shoppers' maximum willingness to pay is $p_{C,k} := q(x_k, k) + \theta\nu(k)$, and quality shoppers are willing

[3] My results are robust to heterogeneity in θ. When the firm faces a single type of conscious shopper, however, it is able in equilibrium to extract all of the surplus from those who buy. With two or more types, the firm is not able to extract all of the surplus by charging one price as long as the firm does not cater exclusively to the most conscious type.

to pay $p_{Q,k} := q(x_k, k)$ for $k = h, m$: Quality shoppers are always willing to pay a weakly higher price for the machine-made product. Depending on the conscious shoppers' concern about the conditions under which the product is manufactured, they may be willing to pay a higher price for the handmade product: $q(x, m) \gtrless q(x, h) + \theta$ for $x > 0$.

In the first stage, the firm infers that quality shoppers are willing to pay less for a handmade product than are conscious shoppers. In particular, the firm sets a price of $p_h^* = p_{Q,h}$ to sell one unit of the handmade product to the whole market. Quality shoppers are indifferent between buying and not buying at p_h^*. The firm sets a price of $p_h^{**} = p_{C,h}$ to sell one unit of the handmade product to a fraction of the market. Conscious shoppers are indifferent at p_h^{**}. Conversely, under Assumption 2, quality shoppers' and conscious shoppers' willingness to pay for a machine-made product coincides. Thus, the firm always sells one unit of the machine-made product to the whole market. The optimal price $p_m^* = p_{Q,m} = p_{C,m}$ makes all shoppers indifferent.

3.3 The Firm's Production Decision

In this section, I analyse the implications of a handmade effect on the firm's optimal production process: Can the presence of conscious shoppers lead to situations in which the firm prefers producing h? Specifically, how strong does the handmade effect have to be to make h preferable?

Proposition 1 (Equilibrium)**.** *If*

$$C\theta > (1 - C)\Pi(p_m^*, x_m^*; m) + C\big[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\big], \tag{3.1}$$

the firm sells h to conscious shoppers only and makes profit

$$\Pi(p_h^{**}, x_h^{**}; h) = C[\Pi(p_h^*, x_h^*; h) + \theta] = C\left[\frac{A_h^2}{2c_h} + \theta\right] > 0,$$

where $p_h^{**} = p_h^* + \theta = A_h x_h^{**} + \theta$ and $x_h^{**} = x_h^* = A_h/c_h$. Otherwise, the firm sells m to the whole market and makes profit

$$\Pi(p_m^*, x_m^*; m) = \frac{A_m^2}{2c_m} > 0,$$

where $p_m^* = A_m x_m^*$ and $x_m^* = A_m/c_m$.

Proof. Detailed proofs of all results are in Section B.2. □

If production is by hand, $k = h$, the conscious shoppers' sensitivity to the production process and the comparative size of segment C determines whether the firm: (i) sells to the whole market at p_h^*; or (ii) sells to conscious shoppers only at a higher price $p_h^* + \theta$. The optimal quality of the handmade product is unaffected by this decision. Moreover, given that profit turns out to be positive for a given production process, the firm always chooses to produce.

Therefore, when conscious shoppers' concern θ is below a certain threshold that decreases in the size of segment C, the firm compares its profit from selling h or m to the whole market. In this case, however, a handmade strategy can never be optimal: The firm cannot charge a handmade premium but faces increased marginal cost and potentially reduced product quality for a given effort level.

For a concern θ above this threshold, the firm compares its profit from selling h to conscious shoppers only with its profit from selling m to the whole market. In this case, a handmade strategy can be optimal if the market segment of conscious shoppers is sufficiently large and concerned enough about the production conditions. This allows the firm to extract a sufficiently high *handmade premium* from this fraction of the market.

In particular, condition (3.1) in Proposition 1 says that the total premium needs to outweigh the loss from: (i) not serving market segment $1-C$; and (ii) selling to segment C under increased marginal cost and reduced product quality for a given effort level. Thus, a handmade product that is sold in equilibrium classifies as a specialised product (Pepall, 1992): It is targeted to those shoppers whose willingness to pay for the special handmade feature is particularly high.

Church and Oakley (2018, p.894) provide empirical support for the handmade premium on Etsy: "Handmade products sold for an average of 78% more than their traditionally manufactured counterparts, even after controlling for shop size, reputation and social media performance". Fuchs et al. (2015) provide experimental evidence: Participants were, on average, willing to pay 17 percent more when a bar of soap was presented as handmade rather than machine-made.

Johnson and Myatt (2006) obtain a similar result: Optimal product design relates to whether a firm uses a mass market or a niche production strategy. They show that when valuations are relatively homogeneous, the firm prefers to serve a large fraction of potential consumers through a product design that has universal appeal. Under heterogeneity, in contrast, the firm restricts sales to a relatively small niche through a product design that caters to specialised tastes. In my model, given Assumption 2, a change from machine to hand production is likewise associated with a dispersion in the distribution of shoppers' willingness to pay, which is profitable for the firm if θ is sufficiently large.

Due to the assumption that $A_m \geqslant A_h \geqslant 1$ and $c_h > c_m > 0$, the firm chooses to produce a higher quality product under machine production. If the firm nonetheless finds it optimal to sell a handmade product to conscious shoppers, regulators may be concerned that the firm substitutes a handmade label for quality, which would leave conscious shoppers ill-informed (Leland, 1979).

Therefore, regulators may introduce a minimum quality standard $\bar{q}$ that will be binding for production by hand, $\bar{q} > q(x_h^{**}, h)$, but will not be binding for machine production, $q(x_m^*, m) > \bar{q}$. In other words, regulation affects only the firm that produces by hand, and forces this firm to increase the handmade product's quality to $\bar{q}$. Under what conditions does such regulation alter the firm's behaviour? Specifically, when does $\bar{q}$ induce the firm to switch to producing m, reversing the handmade effect? How does $\bar{q}$ impact welfare?

Proposition 2 (Minimum Quality Standard)**.** *Suppose condition (3.1) holds: If regulation enforces $\bar{q} \in (q(x_h^{**}, h), q(x_m^*, m))$ and the enforced quality standard exceeds $\check{q}$, where*

$$\check{q} := q(x_h^{**}, h) + \sqrt{q(x_h^{**}, h)^2 - \frac{q(x_m^*, m)q(x_h^{**}, h)}{C} + 2q(x_h^{**}, h)\theta} > q(x_h^{**}, h),$$

then the firm switches to producing m.

Proposition 2 highlights that regulation may have unintended consequences if the enforced quality standard exceeds a certain threshold $\check{q}$, and this threshold falls below $q(x_m^*, m)$. Moreover, a minimum quality standard can alter the firm's behaviour even without productivity differences: The higher effort cost for handmade products induces the firm to choose $q(x_h^{**}, h) < q(x_m^*, m)$ also when $A_h = A_m$. A natural measure of welfare in this setting is the economic surplus from trade: the sum of the firm's profit and the consumer surplus. As the firm increases the price $\bar{p} = \bar{q} + \theta$ one-for-one with increases in the quality standard, the consumer surplus remains constant in $\bar{q}$ and is equal to zero. Therefore, the economic surplus is

$$ES := \begin{cases} C\theta + C\left(\bar{q} - \frac{c_h}{2A_h^2}\bar{q}^2\right) & if \;\; q(x_h^{**}, h) < \bar{q} \leqslant \check{q}, \\ \frac{A_m^2}{2c_m} & if \;\; \check{q} < \bar{q} < q(x_m^*, m). \end{cases}$$

Intuitively, as long as the minimum quality standard does not induce the firm to switch to producing m, a more stringent standard causes the firm's profit, and thus economic surplus, to fall. For a standard that exceeds a certain threshold, the firm switches; and given that the standard is not binding for machine production, economic surplus remains constant. Thus, my result contrasts with the possibility to create welfare-improving standards when there is competition in the dimension of quality (Economides, 1993).

3.3.1 Convex Production Cost

In my benchmark model, I assumed that the per-unit cost remains constant in the level of output produced. I focus on the effect of the production process on optimal effort and product quality, and provide conditions under which the handmade effect induces the firm to produce by hand.

In some settings, however, the workforce may become exhausted when hand production is scaled up due to the manual handling of materials whereas machine production prevents fatigue through mechanisation (ILO, 2010). Also, managerial coordination may become more complex at higher volumes. Thus, I relax the assumption of constant per-unit cost for h in this section. I show that increasing marginal cost reinforces my result that a handmade strategy is associated with lower sales. Moreover, it allows a niche handmade product to be of higher quality than a mass market machine-made product.

I model the firm's profit maximisation problem in such settings as

$$\max_{(p_k,x_k)\in\mathbb{R}^2_+} \Pi(p_k,x_k;k) = \max_{(p_k,x_k)\in\mathbb{R}^2_+} (a_C C + a_Q Q)p_k - (a_C C + a_Q Q)^{\alpha_k}\left[\frac{c_k}{2}x_k^2\right],$$

where $\alpha_h > \alpha_m = 1$. Unlike the benchmark model, the above specification permits: (i) the firm's marginal production cost for the handmade product to be lower than for the machine-made product whenever sales of the handmade product fall short of $s_1 := \left[\frac{1}{\alpha_h}\frac{c_m}{c_h}\left(\frac{x_m}{x_h}\right)^2\right]^{\frac{1}{\alpha_h-1}}$; and (ii) the firm's total production cost for the handmade product to be lower for a given level of output below $s_2 := \alpha_h^{\frac{1}{\alpha_h-1}} s_1$; see Figure 3.2.

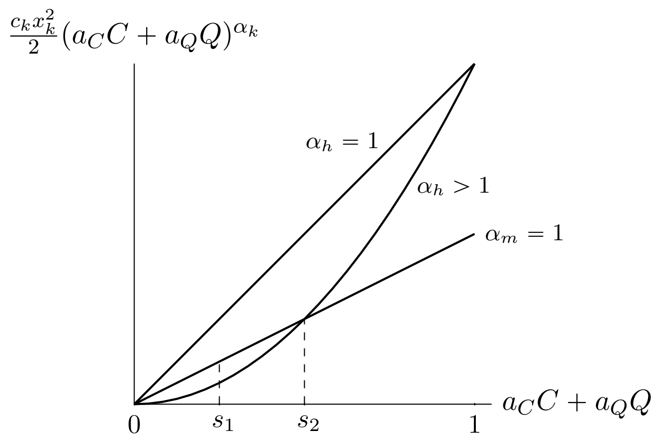


Figure 3.2: Production cost as a function of output.

If production is by hand, $k = h$, the firm's optimal effort x_h^* and profit $\Pi(p_h^*, x_h^*; h)$ remain unchanged when selling the handmade product to the whole market. When selling to conscious shoppers only, the optimal effort level becomes $x_h^{**} = \frac{C}{C^{\alpha_h}}\frac{A_h}{c_h} > x_h^*$, and the firm makes profit $\Pi(p_h^{**}, x_h^{**}; h) = C\left[\frac{C}{C^{\alpha_h}}\frac{A_h^2}{2c_h} + \theta\right] > 0$, where $p_h^{**} = A_h x_h^{**} + \theta$. Therefore, the firm chooses to sell a handmade product to conscious shoppers only at p_h^{**} over the whole market at p_h^* if

$$\theta \geqslant \frac{1 - C^{2-\alpha_h}}{C}\Pi(p_h^*, x_h^*; h). \tag{3.2}$$

Under increasing marginal cost, a niche handmade product has two advantages: First, the firm can charge conscious shoppers a handmade premium. Second, the firm can offer a product of higher quality when selling to conscious shoppers only, rather than to the whole market.

The second effect offsets the decrease in profit from lower sales independent of conscious shoppers' concern: A handmade strategy can be beneficial for the firm even when θ is low: The firm charges a slightly higher price to segment the market, and exclude quality shoppers. This slightly higher price allows the firm to meet demand with a high-quality product. In fact, given $\theta > 0$, the firm will never choose to sell a handmade product to the whole market for $\alpha_h \geqslant 2$: The deterioration in quality and, thus, the reduction in shoppers' willingness to pay would be too large.

As before, a handmade strategy cannot be optimal if condition (3.2) does not hold. If the condition holds, or if $\alpha_h \geqslant 2$, the firm compares its profit from selling a mass market machine-made product with a niche handmade product. In particular, condition (3.1) in Proposition 1 modifies to

$$C\theta > (1-C)\Pi(p_m^*, x_m^*; m) + C\left[\Pi(p_m^*, x_m^*; m) - \frac{C}{C^{\alpha_h}}\Pi(p_h^*, x_h^*; h)\right].$$

The total handmade premium that induces the firm to choose a handmade strategy is lower under increasing than under constant marginal cost. Moreover, the handmade product that is sold in equilibrium is of higher quality than the machine-made product if the fraction of conscious shoppers is sufficiently small: $C < \widetilde{C}$, where $\widetilde{C} := \left(\frac{A_h c_m}{A_m c_h}\right)^{\frac{1}{\alpha_h - 1}} \in (0,1)$. Thus, increasing marginal cost can accommodate handmade products that are associated with higher quality and attention to detail (Dickson and Littrell, 1998).

3.3.2 Multi-Product Monopolist

In my benchmark model, I implicitly assume that the firm is limited to one product for the handmade effect to be credible. When conscious shoppers learn that the firm produces both, they are willing to pay as little as quality shoppers for the handmade product as the firm loses its authenticity. Alternatively, the firm has to commit to having divisions m and h behave as independent profit maximisers to generate the handmade effect (Baye et al., 1996). This divisionalisation induces competition within the firm, and will be analysed in Section 3.4.[4]

[4]The divisionalisation could also be imposed by a regulator due to antitrust concerns. For example, the U.S. Department of Justice judged Grupo Modelo – the producer of the specialty beer "Corona" – to be a "maverick" that is effective at keeping prices low in an oligopolistic market (Davidson, 2013). Therefore, the merger between Modelo and Anheuser-Busch InBev (ABI) was approved conditional on ABI and Modelo divesting Modelo's entire U.S. business to create an independent competitor to ABI (U.S. Department of Justice, 2013).

The restriction that the firm is limited to one product fits, for example, the U.S. craft beer market. U.S. craft breweries are mostly independently owned and are outside larger corporate ownership: "less than 25 percent of the craft brewery is owned or controlled [...] by a beverage alcohol industry member that is not itself a craft brewer" (Brewers Association, n.d.). The world's largest tea company – Unilever – provides an example in the application to Fair Trade (Section 3.6.1): The company has switched to sustainable sourcing of its entire tea supply, including the conversion of its leading tea brand Lipton (Poret, 2010).

There exist, however, counterexamples, such as the banana company Dole, which offers shoppers a standard product and an organic option that is Fair-Trade certified. Thus, I relax the restriction in this section: I show that the unrestricted monopolist always sells to the entire market, and that it adds a second product line h to its standard product line m if the total handmade premium is sufficiently large.

First, note that if the monopolist produces both products and sets prices $p_m^* = q(x_m, m)$ and $p_h^{**} = q(x_h, h) + \theta$, then conscious and quality shoppers are indifferent between buying and not buying m, but only conscious shoppers are indifferent between buying and not buying h. Quality shoppers would not buy h at a premium: $p_{Q,h} < p_h^{**}$. Therefore, offering two product lines with pricing schedule (p_m^*, p_h^{**}) is incentive-compatible for all shoppers as long as I assume that conscious shoppers buy h when indifferent between the two products and not buying.

Moreover, due to the lower marginal cost and the potentially higher product quality under machine production the monopolist always chooses to sell m to quality shoppers to maximise profit. Therefore, the monopolist needs to consider whether it is profit-maximising to add a second product line h in order to sell a handmade instead of a machine-made product to conscious shoppers.

If the firm adds a second product line h, it makes profit $\Pi(p_h^{**}, x_h^{**}; h) = C[\Pi(p_h^*, x_h^*; h) + \theta]$ from selling to segment C; if it offers only one standard product line m to the entire market, the firm makes profit $C\Pi(p_m^*, x_m^*; m)$ from selling to segment C. Comparing the profits in the segment of conscious shoppers, the firm adds product line h if

$$C\theta > C\big[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\big]. \tag{3.3}$$

Compared to condition (3.1), the total handmade premium now only needs to outweigh the loss from selling to segment C under increased marginal cost and reduced product quality for a given effort level. Therefore, the model with a restricted monopolist and an unrestricted monopolist make identical predictions if the total premium is so low that condition (3.3) does not hold: All shoppers are served a machine-made product. If the total premium is moderate so that condition (3.3) but not (3.1) holds, the unrestricted monopolist sells h rather than m to segment C. If the total premium is high, a niche production strategy is optimal only for the restricted monopolist. The unrestricted monopolist offers two product lines to serve the entire market.

In contrast to Mussa and Rosen (1978) and Deneckere and McAfee (1996), the unrestricted monopolist in my model does not cannibalise some of its high-quality market m when introducing the lower-quality product h because my shoppers have identical valuations for quality. Intuitively, the firm does not face a negative externality that would limit its possibilities for capturing all of the consumer surplus when adding a second product line. Hence, the unrestricted monopolist achieves perfect discrimination instead of imperfect discrimination in my model.

3.4 Duopoly

In many product categories – such as soap and ceramic tableware – the availability of high-quality machine production does not prevent firms from selling handmade products (Fuchs et al., 2015): Consumers are given the choice of buying a handmade or a machine-made product. What is more, there is specialisation: The products tend to be sold by different firms. I model this setting with the above framework extended to two firms, F_i and F_j: Each firm can specialise in m or h, and can sell to both segment C and Q. The firms choose the production process, the product's per-unit price, and the per-unit effort level, simultaneously. Shoppers continue to have unit demand.

When is specialisation an equilibrium outcome such that both the handmade product and the machine-made product are offered in the market? Specifically, under which conditions does F_i produce and sell h to conscious shoppers, and F_j produce and sell m to quality shoppers? What are the equilibrium prices and profits under specialisation?

Proposition 3 (Specialised Equilibrium)**.** *There exists an equilibrium in which F_i sells h to conscious shoppers at $p_h^{**} - \bar{\varepsilon}$, and F_j sells m to quality shoppers at p_m^* if*

$$\theta \geqslant \left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) \geqslant \left(\frac{C}{1-C} + \frac{1-C}{C}\right)\Pi(p_h^*, x_h^*; h). \quad (3.4)$$

Under specialisation, F_i and F_j make the same profit and the price discount offered on h is

$$\bar{\varepsilon} = \theta - \left[\left(\frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\right] < \theta.$$

Specialisation is an equilibrium outcome if the conscious shoppers' sensitivity to the production process is sufficiently high. Specifically, condition (3.4) prevents a profitable deviation for F_j and F_i, respectively.[5] Only a one-sided price reduction on h but not m can be a specialised equilibrium: F_j charges p_m^*, while F_i charges $p_h^{**} - \bar{\varepsilon}$. Given Assumption 2, conscious shoppers would want to deviate to buying a machine-made product if F_j reduced p_m^* while F_i charged p_h^{**}.[6] The discount offered on h is strictly positive and bounded above by θ, which prevents a profitable deviation for quality shoppers. The price discount on h leaves conscious shoppers with strictly positive utility, whereas quality shoppers are left with zero utility: They are indifferent at $p_m^* = p_{C,m}$.

Proposition 3 highlights that the presence of conscious shoppers and the choice of production process $k \in \{h, m\}$ can break the usual price competition. Instead of a Bertrand duopoly in which the two firms target overlapping market segments and compete prices down to marginal cost to capture shoppers, F_i and F_j choose to act as monopolists in distinct market segments. The firms can act as monopolists because the segments do not adjust in response to relative price changes.

My result is in line with d'Aspremont et al. (1979, p.1149), who note that "oligopolists should gain an advantage by dividing the market into submarkets in each of which some degree of monopoly would reappear". A similar result has been obtained by Shaked and Sutton (1982) in a different setting: When consumers are identical in tastes but differ in income, price competition is relaxed through product differentiation. Mohliver et al. (2019) obtain differentiation in a counter-positioning equilibrium. Firms take opposing stances on a social issue that is characterised by: (i) how much people care about the issue; and (ii) what proportion of the population supports the issue.

The intuition behind Proposition 3 is akin to the one of Mussa and Rosen (1978): A sufficiently strong sensitivity allows the two firms to choose a price-quality schedule that allocates the two shopper types to the two different products h and m by a process of self-selection. Furthermore, given that conscious shoppers focus more on production conditions than on quality, tastes have to be sufficiently diverse for the handmade product to co-exist alongside the machine-made product of higher quality (Eaton and Lipsey, 1989).

An important question for policy is whether such specialisation is efficient. I define an outcome as efficient if it maximises the economic surplus from trade between the firms and the shoppers. The surplus is the sum of those shoppers' marginal benefits that buy a product in equilibrium minus the production costs; equivalently, this is the sum of the firms' profits and the consumer surplus. Therefore, the economic surplus in the specialised

[5] If condition (3.4) does not hold, specialisation with a two-sided price reduction can be an equilibrium. The specialised equilibrium in Proposition 3 is robust to k being chosen before p_k and x_k. The results are available in Section A.2.1.

[6] If instead $\nu(m) < 0$, a one-sided price reduction $\hat{\varepsilon} \in (0, \nu(m)\theta]$ on m would be incentive-compatible for C.

equilibrium is

$$ESS := QA_m x_m^* + C(A_h x_h^* + \theta) - Qc(x_m^*) - Cc(x_h^*)$$
$$= Q\Pi(p_m^*, x_m^*; m) + C\Pi(p_h^*, x_h^*; h) + C\theta = 2Q\Pi(p_m^*, x_m^*; m) + C\bar{\varepsilon}.$$

Given that conscious shoppers obtain the handmade product and quality shoppers obtain the machine-made product in the specialised equilibrium, a scenario in which the two firms compete in selling h to conscious shoppers only cannot be more efficient. Furthermore, it follows from Proposition 1 that competing in selling h to the whole market cannot be more efficient either. Therefore, I can focus on F_i and F_j competing in selling m to the whole market to establish the specialised equilibrium's efficiency.

Proposition 4 (Efficiency)**.** *The specialised equilibrium is efficient.*

Intuitively, the existence of a specialised equilibrium implies its efficiency: The conscious shoppers' sensitivity to the production process and its associated surplus $C\theta$ is so high in the specialised equilibrium that it trumps the loss $C[\Pi(p_m^*, x_m^*, m) - \Pi(p_h^*, x_h^*; h)]$ in surplus from selling a lower-quality product to conscious shoppers.

Some of the gains from efficiency are passed along to conscious shoppers because F_i's pricing behaviour is disciplined by the presence of a potential competitor (Bain, 1949): The discount $\bar{\varepsilon}$ serves as a means to deter the switching of F_j into hand production, and thereby coordinates the firms' behaviour on specialisation. Consequently, the potential competition moderates F_i's profit but falls short of more vigorous competition (Gilbert, 1989): The discount is smaller than the conscious shoppers' surplus that would be created by perfect competition.

In my analysis so far, I have made the implicit assumption that F_i and F_j play non-cooperatively. I now relax this assumption and allow the two firms to collude in order to study the robustness of the specialised equilibrium in Proposition 3. Under which conditions do the firms not have an incentive to merge to a monopolist, selling a single product k, when sharing the monopolist's profit equally? In particular, either F_i can merge with F_j to sell m to the whole market, or F_j can merge with F_i to sell m to a fraction of the market.

Proposition 5 (Stability)**.** *The specialised equilibrium is stable if* $\theta \leqslant \frac{2(1-C)}{C}\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)$ *and* $C \leqslant \frac{1}{2}$.

Corollary 1 (Collusion)**.** *In the specialised equilibrium, there exists* $\bar{C} := \frac{\sqrt{5}-1}{2} > \frac{1}{2}$ *such that* F_i *and* F_j *can profitably collude as (i) h- or m-monopolist for* $C > \bar{C}$*, and (ii) m-monopolist for* $\bar{C} \geqslant C > \frac{1}{2}$.

Intuitively, the specialised equilibrium is robust to collusion if the proportion of conscious shoppers in the market is not too large and the conscious shoppers' concern about the conditions under which the product has been manufactured is not too high. Specifically, merging to sell m to the whole market is profitable for F_i and F_j when conscious shoppers are in the majority. Merging to sell h to a fraction of the market requires a larger critical mass $\bar{C}$.

3.5 Signaling a Social Norm of Conscious Consumption

The baseline model with conscious and quality shoppers explains why a firm chooses production by hand over machine production through a handmade effect on the consumer side. However, the baseline model does not capture the rising demand for and the mainstreaming of costly handmade products (Richards, 2018), the sales growth on dedicated marketplaces like Etsy (Figure 3.1), and social movements such as "buyhandmade.org" (Walker, 2007).

In this section, I consider how the firm's handmade strategy can be indicative of a social norm of conscious consumption. As a consequence, shoppers who are not conscious to begin with convert into conscious shoppers and also buy premium-priced handmade products. I develop a signaling model with a single firm in the spirit of Sliwka (2007): quality and conscious shoppers' decision criteria are fixed from the outset, whereas conformist shoppers are influenced by social norms. Under which conditions can the firm credibly signal a social norm of conscious consumption, inducing those conformist shoppers to buy premium-priced handmade products in equilibrium?

3.5.1 The Extended Model

Shoppers come in three types: As before, quality shoppers care only about product quality whereas conscious shoppers also care about the conditions under which the product is produced. Following Sliwka (2007), I refer to these two types as the *steadfast* shoppers: Their decision criteria are fixed from the outset. The fraction of conscious shoppers among the steadfast is given by φ, and is drawn from some prior distribution.

Conformist shoppers constitute the third type. Because these shoppers are uncertain about the appropriate consumption behaviour, they are influenced by social norms. For example, if the firm is selling a handmade product, conformist shoppers will take the production conditions positively into account only if they believe that many other shoppers also do. I model this as follows: The production process is payoff-relevant for a conformist shopper if and only if she believes that sufficiently many other steadfast shoppers are also conscious about the production process. In particular, the utility of a conformist shopper is equal to $V_C(a_C;k)$ if she believes that the median steadfast shopper is conscious, and equal to $V_Q(a_Q;k)$ otherwise.[7]

As an established firm, it will typically have learned from previous shoppers' behaviour or will have undertaken market research. Therefore, I assume that the firm knows the fraction of conscious shoppers among the steadfast to be either high or low, $\varphi \in \{\varphi_H, \varphi_L\}$. Moreover, I assume that the firm's private information about the fraction of conscious shoppers is pivotal: $\varphi_H > 1/2 > \varphi_L$.

Consequently, when conformist shoppers are able to infer the firm's private information from its production process, they revise their preferences. Conversely, when conformist shoppers cannot infer the firm's private information, then $\mathbb{E}(\varphi) = \gamma\varphi_H + (1-\gamma)\varphi_L \lesseqgtr 1/2$ with commonly known prior $\gamma := \Pr(\varphi = \varphi_H) \in (0,1)$ determines whether conformist shoppers behave like conscious or quality shoppers. The fraction of conformist shoppers in the market has mean η according to the common prior expectation; see Figure 3.3.

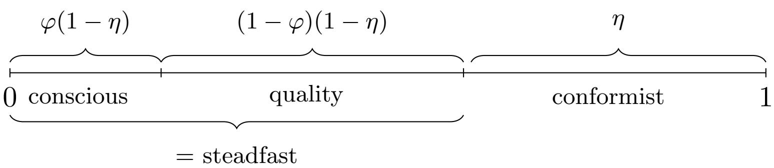


Figure 3.3: The extended market.

The timing is as follows. In the first stage, the firm privately observes $\varphi \in \{\varphi_H, \varphi_L\}$. As before, the firm chooses the production process $k \in \{h, m\}$, the product's per-unit price p_k, and the per-unit effort level $x_k \geqslant 0$. In the second stage, the shoppers choose whether to buy one unit of the product. All shoppers have unit demand irrespective of the production process.

Given that the firm's choice of the production process may reveal its private information, the extended model is a signaling game. Moreover, depending on a conformist shopper's belief about φ, her action will always correspond exactly to either that of a steadfastly quality shopper or a steadfastly conscious shopper. Hence, the firm's profit in the extended model depends on its production process k and whether conformist shoppers act like conscious or quality shoppers.

Assumption 3. $\Pi(p_h^*, x_h^*; h) + \theta > \Pi(p_m^*, x_m^*; m)$.

[7]Conformist shoppers may pay attention only to steadfast shoppers if the latter segment consists of public persona. These steadfast shoppers are visible through the news (e.g., celebrities) or social media (e.g., bloggers, influencers). In an alternative version, a conformist shopper notices anyone *behaving* like a conscious shopper. If I assume that the firm's private information is important to coordinate conformist shoppers to convert, $\varphi_H > (1-2\eta)/(2-2\eta) > \varphi_L$, where $\eta < 1/2$, Proposition 6 still applies. However, for $\varphi_H < 1/(2-2\eta)$, the separating equilibrium is "less stable", since a conformist shopper is only indifferent between buying and not buying h at a premium if her segment buys h, too.

Before deriving the key result, I make the assumption that the firm would be better off using a handmade strategy if the market consisted entirely of steadfastly conscious shoppers.

3.5.2 Separating Equilibrium

The key idea is that it may be attractive for the firm to choose a handmade strategy to signal its private information that most shoppers are conscious. However, even a credible handmade strategy comes with a trade-off: While the firm is better off when conformist shoppers convert into conscious shoppers, quality shoppers do not buy a premium-priced handmade product.

The handmade strategy will be a credible signal when there is separation in equilibrium: A firm that privately observes φ_H produces by hand, and a firm that privately observes φ_L uses machine production. Specifically, when the conformist shoppers believe that the handmade strategy is indeed a credible signal, neither firm can have an incentive to deviate from this strategy. If φ is the firm's private information about the fraction of conscious shoppers among the steadfast, its expected profit under machine production is given by

$$\Pi(p_m^*, x_m^*; m) = [\varphi(1-\eta) + (1-\varphi)(1-\eta) + \eta]\frac{A_m^2}{2c_m} = \frac{A_m^2}{2c_m}. \tag{3.5}$$

Given Assumption 2, the firm sells to the whole market at price $p_m^* = A_m x_m^*$. Intuitively, under a one-sided concern of conscious shoppers, the market composition is irrelevant for the firm's pricing decision under machine production.

When the firm produces by hand, it must prefer losing out on the fraction of quality shoppers. This follows from Proposition 1: A handmade strategy can never be optimal for a firm that sells to the whole market of shoppers. The firm would have an incentive to deviate. Hence, I assume that $\eta \geqslant \tilde{\eta}(\varphi_H)$ in the separating equilibrium, where

$$\tilde{\eta}(\varphi_H) := \frac{1}{1-\varphi_H}\left[\frac{\Pi(p_h^*, x_h^*; h)}{\Pi(p_h^*, x_h^*; h) + \theta} - \varphi_H\right],$$

which implies that $p_h^{**} = A_h x_h^* + \theta$ and $x_h^{**} = x_h^* = A_h/c_h$ are optimal. As quality shoppers' maximum willingness to pay is $p_h^* = A_h x_h^*$, this type will not buy the handmade product.

Simultaneously, the firm benefits from an *amplified handmade effect*: Because conformist shoppers become conscious, they are willing to pay the handmade premium. Hence, there is greater demand for the premium-priced handmade product. The firm's expected profit under hand production is given by

$$\Pi(p_h^{**}, x_h^{**}; h) = [\varphi(1-\eta) + \eta]\left[\frac{A_h^2}{2c_h} + \theta\right]. \tag{3.6}$$

Comparing equation (3.5) and equation (3.6), and solving for η highlights that the firm will use a handmade strategy when the fraction of conformist shoppers is larger than the threshold

$$\hat{\eta}(\varphi) := \frac{1}{1-\varphi}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h) + \theta} - \varphi\right].$$

This threshold is decreasing in φ, since higher values imply a lower fraction of steadfastly quality shoppers who do not buy the premium-priced handmade product. As a result, a handmade strategy becomes more attractive. I use the threshold to derive the separating equilibrium.[8]

Proposition 6 (Separating Equilibrium). *If $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$, a separating equilibrium exists in which the firm sells a premium-priced handmade product to steadfastly conscious and conformist shoppers after observing φ_H, and sells a machine-made product to the whole market after observing φ_L if and only if*

$$\eta \in [\max\{0, \hat{\eta}(\varphi_H)\}, \hat{\eta}(\varphi_L)],$$

where $\hat{\eta}(\varphi_H) < \hat{\eta}(\varphi_L) < 1$ and $\hat{\eta}(\varphi_L) > 0$.

The conscious shoppers' sensitivity to the production process and the fraction of conformist shoppers jointly determine whether the separating equilibrium in Proposition 6 exists. The greater is the fraction of conformists, the smaller is the range of concerns θ for which separation is possible: A handmade strategy becomes more appealing because the firm's expanded market for a premium-priced handmade product grows faster than its base market shrinks.

Moreover, if θ is small, a separating equilibrium exists when there are neither too few or too many conformists. Specifically, when $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_H - \Pi(p_h^*, x_h^*; h)$, the lower bound is given by $\hat{\eta}(\varphi_H)$. When there are few conformist shoppers, even a firm that privately observes φ_H prefers to sell a machine-made product. The higher is the fraction of conformist shoppers, the more attractive it is to signal optimism about the social norm of conscious consumption. But when there are too many conformist shoppers, even a firm that privately observes φ_L would want to imitate this signal, and it would no longer be credible.

If the conscious shoppers' sensitivity is above this value, the total handmade premium from selling to steadfastly conscious and conformist shoppers is so high that a firm that

[8]There can be other pure-strategy equilibria where the φ_L- and φ_H-firm pool on selling h or m, and a hybrid equilibrium where the φ_H-firm sells h with certainty and the φ_L-firm sells h with probability $x \in (0, 1)$. However, separation is the unique pure-strategy equilibrium in the parameter space of Proposition 6. Moreover, the φ_L-firm pools more often with the φ_H-firm on h in this game of private information relative to when φ is observable. My analysis extends to a situation in which conformists' willingness to pay is increasing smoothly in $\mathbb{E}(\varphi|k)$. The results are available in Section A.2.2.

observes φ_H would never want to deviate to machine production. In this case, the separating equilibrium exists when there are not too many conformists and $\theta < \Pi(p^*_m, x^*_m; m)/\varphi_L - \Pi(p^*_h, x^*_h; h)$.

Intuitively, the fewer are the number of steadfastly conscious shoppers, the higher is the cost of using a handmade strategy in the absence of conformists: The firm cannot offset the loss in sales to quality shoppers through higher prices that are charged to conformists. Hence, a firm that knows that the fraction of conscious shoppers among the steadfast is low may want to avoid a handmade strategy even if such a strategy could increase the conformists' willingness to pay. In other words, separation is possible when this firm does not find it profitable to convert conformists as there are too few.[9]

The key result in Proposition 6 is that the firm can use a handmade strategy to create a credible social norm of conscious consumption: The firm that observes φ_H is able to expand its market and mainstream costly handmade products. In essence, a credible signal generates sales growth through a *crowding-in effect*: Conformist shoppers convert into conscious shoppers when observing the firm's handmade strategy. Importantly, this crowding-in effect can operate in a market in which the firm that observes φ_H believes that steadfastly conscious shoppers are in the minority. Specifically, even when the fraction of steadfastly conscious shoppers is arbitrarily small (i.e. $\eta \to 1$), a handmade strategy may credibly signal a social norm of conscious consumption.

Norm-based interventions and marketing campaigns (e.g., Cialdini et al., 2006) provide empirical support for my result. The credibility problem that limits a campaign's effectiveness is highlighted by Benabou and Tirole (2011). The interaction between conscious shoppers' concern and credibility is also observed by Melloni et al. (2019): In a cheap-talk model, CEO activism is more likely to be a credible and, thus, profitable strategy when shoppers care greatly about the type of firm that they buy from.

3.6 Discussion

In this paper, I developed a model that explains a handmade strategy of firms in an era of technological advancement through a handmade effect on the consumer side.

3.6.1 Further Applications: Fair Trade

My model generates insights whenever a firm chooses between a conventional production process and an alternative production process. The latter includes Fair Trade, organic produce, and local produce. To demonstrate, I relate my assumptions to Fair-Trade production and underline my predictions with empirical evidence in Fair Trade.

[9] I thank an anonymous referee for inspiring this intuition.

My assumptions capture several key characteristics of Fair Trade: First, Fair-Trade products and conventional products – such as coffee and chocolate – are functional substitutes for shoppers: "Fair Trade and traditionally marketed products show at most very little divergence and hardly differ with respect to their functional utility" (Mohan, 2010, p.34). Second, Fair Trade may be less productive in the short-run. While higher environmental standards ensure sustainability, land may not be farmed as extensively. Farmers are also encouraged to reduce their use of pesticides (Nicholson, 2018). Fair Trade is more costly for the same per-unit effort level: Firms pay a licence fee for the Fair-Trade label, and pay a living wage to workers, which may exceed the legal minimum.[10] Moreover, providing a safe working environment may increase costs.

Conscious shoppers obtain utility from buying a product with a Fair-Trade label.[11] Importantly, these shoppers "care about the production process not because of its effects on the quality of the product itself but for other attributes of the process, such as environmental impacts, effects on animal welfare, country of origin, and so on" (Antle, 1999, p.1002). In essence, the Fair-Trade firm sells a bundle: The conventional product, plus a contribution to a social or ecological production process (Steinrücken and Jaenichen, 2007) or a donation to the farmer (Reinstein and Song, 2012). As is argued by Richardson and Stähler (2014), the mechanism may be a warm glow from the knowledge that a Fair-Trade firm pays its growers a wage premium.

In my baseline model, selling a Fair-Trade product is the outcome of a profit maximisation problem: The firm weighs the additional costs against the expected benefits. In particular, the firm faces the trade-off that it can charge a Fair-Trade premium but only caters to a niche market of conscious shoppers. In my model, this Fair-Trade premium is the conscious shoppers' maximum willingness to pay for the subjective services of the Fair-Trade label.[12] My model is in line with the observation that Fair-Trade products are more expensive than conventional products but do not deliver any extra physical quality (Mohan, 2010).

When the firm is part of a duopoly, conventional and Fair-Trade products may co-exist, as is observed in the economy. Moreover, the firms' specialisation leads to market segmentation and allows for third-degree price discrimination: The Fair-Trade label prevents

[10] A study of Fair-Trade workers in the South African wine industry, for example, finds that wages exceed the legal minimum (Granville and Telford, 2013). Moreover, wages in Fair Trade-certified firms in Ghana are about 38 percent higher than wages in non-certified firms (Krumbiegel et al., 2018).

[11] Using data from the Understanding Society UK survey, Busic-Sontic et al. (2017) show that *conscientiousness*, which could be one possible motivation for conscious shoppers, positively affects buying locally sourced goods, organic or green products, and products with recyclable packaging.

[12] See Dragusanu et al. (2014) for a literature review on how much consumers are willing to pay for knowing that a product's production complies with Fair-Trade standards. See Bjørner et al. (2004) for environmental labels.

conventional products from entering the Fair-Trade market. When the firm faces conformist shoppers, my model predicts a growth in Fair-Trade sales if the firm can use a Fair-Trade strategy to create a credible social norm of conscious consumption. In particular, the conversion of conformist shoppers into conscious shoppers resembles the increased public awareness and understanding of the Fair-Trade rationale, and its mainstreaming in Europe in the late 1990s (Renard, 2003).[13]

The predicted crowding-in effect aligns with the finding that building positive attitudes and removing skepticism positively impacts Fair-Trade buying behaviour (De Pelsmacker and Janssens, 2007). Consequently, my model not only predicts that "consumers of Fair Trade products form a kind of speciality market club, which can be joined voluntarily by those who are willing to pay extra for the Fair Trade product bundle" (Mohan, 2010, p.43), but also that the sales growth of Fair-Trade products in Europe can be explained through a credible signal that most steadfast shoppers care.

3.6.2 Concluding Remarks

In my analysis, I made several simplifying assumptions: While a one-sided concern of conscious shoppers kept my model tractable, one could envisage situations in which regulation enforces firms to state their production process on the packaging. As a result, both hand and machine production are salient when conscious shoppers make the buying decision: $\nu(h) > 0 > \nu(m)$. A firm that chooses machine production may then not want to serve all shoppers because conscious shoppers' maximum willingness to pay falls short of quality shoppers' maximum willingness to pay: $p_{C,m} < p_{Q,m}$.

I also assumed that conscious shoppers and quality shoppers do not interact. However, when quality shoppers buy a handmade product that is also purchased by conscious shoppers, they may incur positive spillovers. For example, quality shoppers may value being perceived as conforming to a social norm of green consumerism (Carlsson et al., 2010). Future research could adapt the baseline model to a setting in which quality shoppers are willing to pay more than what is warranted by the handmade product's quality if they are likely to meet a conscious shopper who bought the handmade product in stage 3.[14] Consequently, a handmade strategy may also be optimal for a firm that caters to the whole market.

[13] See Friedrichsen and Engelmann (2018) for experimental evidence that consumers who choose a conventional product in private increase the premium that they are willing to pay for a Fair-Trade product in a public setting.

[14] See Sebald and Vikander (2019) for a model where shoppers are matched pairwise after the buying decision, and where the peer's beliefs about the purchased product's popularity affects a shopper's social image.

Finally, future work could endogenise conscious shoppers' sensitivity to the production process through a handmade spectrum (Etsy, Inc., 2019b): At one end would be producers literally making products with their own hands; and producers designing products with partners physically producing them would fall into an intermediate range. Instead of a binary choice k of the production process, a firm would then choose $\theta \in [0, \bar{\theta}]$ to position itself on the spectrum, where $\theta = 0$ corresponds to the machine production of all components and per-unit cost $c(x_\theta)$ increases in θ.

Chapter 4

Conform or Reveal? The Effect of Heuristics on Behaviour*

4.1 Introduction

The Peanuts comic strip originally presented here cannot be made freely available via ORA because of third-party copyright. The comic strip from 25 April 1989 is reproduced in The Complete Peanuts: 1989 to 1990 By Charles M. Schulz, Volume 20, page 50 (centre panel).

Figure 4.1: Lucy and Linus in Peanuts (25 April 1989; United Feature Syndicate, Inc).

In the syndicated comic strip *Peanuts*, Lucy van Pelt is quick to make trait attributions when asked about the behaviour of her younger brother Linus; yet, when reflecting upon her own behaviour, she is focused on the situation (Figure 4.1). The comic strip summarises a long-known insight from psychology: individuals tend to attribute their own behaviour to situational causes but explain the identical behaviour of others through their inherent qualities or traits (Jones and Nisbett, 1971; Nisbett et al., 1973).[1] Despite using others' actions to guide their decisions, individuals often overlook that others also examine the social evidence (Cialdini, 1984). Under uncertainty, such trait attributions serve as a *heuristic* because they allow individuals to reduce the difficulty of their decisions (Tversky and Kahneman, 1974).

In this paper, we study the implications of the above heuristic when individuals (i) compare their behaviour with others', but (ii) have no means of communicating with them.

*Co-authored with Dr Matthew J. Robertson.

[1]One reason may be that individuals observe others but not themselves (Taylor and Fiske, 1975). Trait attributions may also be easier to make than situational attributions (Newman and Uleman, 1989), and the former may be needed to make sense of an event (Read, 1987).

We set out a model in which players face a trade-off between conforming to the average action taken by others and matching their own type. A player's type is private information and the unknown state of the world determines which type is more likely held by the majority of players. After being sorted into a random order, each player chooses a publicly observable action given her type and the history of actions.

We contrast individual and aggregate behaviour under two heuristics and rationality. A *credulous* player posits that she alone is influenced by social comparison: she presumes that every other player truthfully reveals their type. Similarly, a *sceptical* player posits that each of her successors will truthfully reveal; however, she reflects on the salient actions of her predecessors and discounts those for which there was a majority at the time the action was taken. A fully *rational* player takes into account the extent to which others wish to conform to the group's behaviour.

We begin our analysis of individual behaviour with a characterisation of the conditions under which a player using a heuristic falsifies her type, and how her decision is affected by the number of predecessors in agreement with her type. We use this result to understand how the two heuristics affect individual behaviour: when do sceptical players behave differently from credulous players? To build towards our analysis of aggregate behaviour, we study whether, compared to heuristic decision-making, rationality reduces a player's incentive to falsify. We then answer our main research question on aggregate behaviour: does a more complex heuristic make an outcome in which everyone agrees on a certain action less likely?

We contribute several insights for policy. In our analysis of individual behaviour, we show that a policy maker can *boost* a credulous player to scepticism if his goal is to increase truth telling.[2] In contrast, a boost to rationality does not guarantee that more players truthfully reveal. An intervention that lowers the pressure to conform in the presence of heuristic decision-making may be more effective. In our analysis of aggregate behaviour, we show that a boost to rationality is not only necessary to incentivise players to take the action matching the state irrespective of their types, but also preferable given a utilitarian objective. Finally, making the state of the world common knowledge can be a viable alternative when a policy maker prefers a player to tell the truth if her type matches the state and to falsify if it does not.[3]

[2]Boosts develop an individual's competences through decision tools like procedural routines or decision trees to enable specific behaviours (Hertwig and Grüne-Yanoff, 2017). Nudges, in contrast, target behaviour directly through a change in the choice architecture.

[3]For intuition how a policy maker may know the state and can make it common knowledge, consider ideological voting. In this setting, the state of the world determines which of two candidates is more factual; a player's type is her political ideology, which is more likely aligned with the more factual candidate. The player's objective is not to vote for the more factual candidate but to hold true to her ideology. The policy maker is an independent watchdog who knows the state of the world by fact-checking candidates' election

We apply our insights to organisational decision-making and social engineering. We consider how the desire for conformity may lead a board or committee to reach a consensus at all costs without critical appraisal of the alternatives.[4] We then discuss how this tendency to agree may be reduced by (i) encouraging credulous players to reflect on the history of actions in a more critical way, or (ii) reducing individuals' desire to conform. In the latter application, we discuss how a policy maker can incentivise a re-thinking on important issues such as climate change by (i) reducing the uncertainty regarding the majority opinion, (ii) revealing the state of the world, or (iii) making credulous players rational.

4.1.1 Related Literature

Our model draws upon literature in social psychology: there is extensive experimental evidence that individuals underestimate the influence of social pressure on people's public actions. For example, in Miller et al.'s (1974) adaption of the Milgram (1963) experiment, subjects witnessed teachers obeying instructions to deliver electric shocks to individuals and were then put in the role of teachers themselves. Yet, when predicting how other teachers would behave based on these two experiences, subjects heavily underestimated the influence of the social environment. We take these insights and model this underestimation as a consequence of using a heuristic to simplify a decision problem under uncertainty.

We also build on empirical evidence in economics that a lack of communication can sustain incorrect judgments about others' beliefs. Bursztyn et al. (2020), for example, find that most Saudi men privately support that women should be allowed to work outside the home (WWOH) but substantially underestimate the proportion of men sharing their view. What is more, men whose male friends and relatives discuss WWOH rarely or very rarely underestimate the level of support for WWOH by a significantly greater amount (40 percent) than men who communicate often or very often (4 percent).

Our players' preferences are similar to those in Kuran (1987a,b, 1995). A player may suffer a loss in utility when deviating from the arithmetic mean of others' actions. We abstract from the exact mechanism underlying this loss, such as disapproving gestures of others, being denied opportunities, or losing one's social status (Kuran, 1995). Moreover, a player may have a need for self-assertion: she may suffer a loss in utility from not matching her own private opinion in public. Given these two objectives, our players face a similar trade-off between social approval and personal autonomy: conforming may preclude truthfully revealing one's type. By adapting Kuran's (1995) concept of preference falsification, we derive a condition under which a player chooses an action that differs from her type. In

programmes. The watchdog can make the state common knowledge by disseminating information on which programme makes more scientifically sound claims.

[4] This phenomenon is known as groupthink and is discussed in the context of our model in Section 4.6.1.

Kuran (1987a,b), however, players move simultaneously. The resulting informational symmetry justifies his assumption that all players have the same point estimates of the share of players taking each of the two actions. In our model, in contrast, players move sequentially, gaining additional information as the history evolves. This allows us to study how heuristic decision-making interacts with an informational asymmetry to trigger preference falsification.

The hierarchical model of bounded rationality in Stahl and Wilson (1995) is similar in that each player type is incrementally more rational: level-0 resemble credulous, and level-1 resemble sceptical players who think everyone else is credulous. A worldly player resembles a sceptical player in the alternative specification of Section A.3.2 as she believes she faces a naive type with probability ε, and effectively a level-1 type with probability $1-\varepsilon$. However, Stahl and Wilson (1995) assume that all types compute their expected utility with error to test the theory experimentally.

Our discussion of (i) organisational decision-making and (ii) social engineering contributes to the literature on interventions to mitigate conformist groups' tendency to agree (e.g., Benabou, 2013) and incentivising political correctness (e.g., Morris, 2001), respectively.

In Section 4.2, we introduce the model. In Section 4.3 and Section 4.4, we compare individual and aggregate behaviour under two heuristics as well as rationality. In Section 4.5, we analyse the implications of changes to the informational environment. We conclude with two applications in Section 4.6. All proofs are in Section B.3.

4.2 Model

Consider a set $\mathcal{N} := \{1, \ldots, N\}$ of players who interact over N periods. Let i and j denote generic players in $\mathcal{N}$. In period -1, the uniformly distributed state of the world $\omega \in \{0, 1\}$ is realised. In period 0, the players are randomly indexed from 1 to N. A player's index determines the period in which she moves; for example, player $i = 3$ moves in period 3. Each player also privately learns her type $\theta_i \in \{0, 1\}$. The players' types are independent, identically distributed and linked to the state by the conditional probability $\Pr(\theta_i = 0|\omega = 0) = \Pr(\theta_i = 1|\omega = 1) := \pi \in \left(\frac{1}{2}, 1\right)$. Consequently, π measures the informativeness of a player's type about the state, and ω is more likely to produce groups with majority type $\theta = \omega$. Both the distribution of the state and the parameter π are common knowledge.

In periods 1 to N, each player takes an action that is publicly observable. Player i chooses $s_i \in \{0, 1\}$ in period i after observing her type θ_i and the history $h_i = \{s_1, \ldots, s_{i-1}\}$

to solve $\max_{s_i \in \{0,1\}} \mathbb{E}[U(s_i; \theta_i, \lambda_i)] =$

$$\max_{s_i \in \{0,1\}} -(1-\lambda_i)\mathbb{E}\big[(s_i - \bar{s}_{-i})^2 | h_i, \theta_i, s_i\big] - \lambda_i (s_i - \theta_i)^2, \tag{4.1}$$

where $\lambda_i \in [0,1]$ and $\bar{s}_{-i} := \frac{1}{N-1}\sum_{j \neq i} s_j$. Each player is concerned with fitting in: she prefers to be viewed as conforming to the average action in the group. Being perceived as deviant by the other group members is costly. However, each player also prefers to choose an action that corresponds to her type: dissonance between one's action and type is costly. The weight λ_i determines the importance of conforming relative to matching one's type.

We assume that a player does not care about the state of the world per se. This assumption is appropriate in settings where a single player is too inconsequential to have an impact (Downs, 1957). Kahan (2013, p.409) provides climate change as an example of such a setting: amid scientific consensus on the reality of climate change, an individual's denial will not affect her utility. What matters for her is whether her position on climate change disagrees with the predominant position in her social group as this may compromise her relationships or status and, thus, her welfare.[5]

4.2.1 Players' Presumptions

Let $\hat{\lambda}_{ij}$ denote player $i \neq j$'s presumption about player j's conformity motive, where $\hat{\lambda}_{ij}$ need not coincide with player j's true conformity motive λ_j.

Definition 1 (Credulity)**.** Player i is credulous if $\hat{\lambda}_{ij} = 1$ for all $j \neq i \in \mathcal{N}$.

A *credulous* player posits that every other player reveals her type. Therefore, she also presumes that every other player believes that she reveals her type. In reality, however, each player's action is determined by both her type and the actions of others. Consequently, from a credulous player i's perspective, each player $j \neq i$ solves $\max_{s_j \in \{0,1\}} U(s_j; \theta_j, 1)$, which has the unique solution $s_j^*(\theta_j) = \theta_j$.

Given Definition 1, a credulous player can simplify her inference problem. She presumes that her action does not affect the actions of the other players.[6] Thus, she can use a summary statistic $M_i := \sum_{j \in h_i} \mathbb{I}\{s_j = \theta_i\}$ that keeps track of the number of predecessors

[5] The debate can continue in public discourse amid scientific consensus because faith in the truth (of global warming) is undermined for ideological reasons by a small number of contrarian "experts" (Oreskes and Conway, 2010).

[6] A credulous player provides the same answer to the questions "How will my community think of me if I fail to endorse its dominant position?" and "What is the dominant view within my society?" (Kuran and Sunstein, 1999), thereby blending a reputational motive (e.g., Bernheim, 1994; Kuran, 1995) with an informational motive (e.g., Banerjee, 1992; Bikhchandani et al., 1992). Hence, when players use heuristics, these two drivers of conformist behaviour identified in the economics literature can overlap.

that have publicly agreed with her type, where $M_1 := 0$. Hence, the decision to truthfully reveal, $\mathbb{E}[U(\theta_i;\theta_i,\lambda_i)] \geqslant \mathbb{E}[U(1-\theta_i;\theta_i,\lambda_i)]$ reduces to $\lambda_i \geqslant \lambda_i(M_i,\theta_i)$,[7] where

$$\lambda_i(M_i,\theta_i) := 1 - \frac{1}{2\big(\theta_i + (1-2\theta_i)\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]\big)}.$$

Our motivation for Definition 1 is twofold. First, individuals often use a mental shortcut to reduce the difficulty of decisions under uncertainty (Tversky and Kahneman, 1974). The availability heuristic, in particular, may induce a player to solve a simplified version of (4.1): because the history of actions is readily recalled when having to forecast successors' actions, a credulous player i acts on the notion that h_i must be important to the task at hand (Esgate and Groome, 2005). Without other readily accessible information, she is led to make the simplifying assumption in Definition 1 to apply a straightforward forecasting rule that uses the distribution of actions in h_i as a predictor of the distribution of the following $N-i$ actions. Second, when confronted with uncertainty, individuals tend to use others' actions to decide how they themselves should behave. Yet, in so doing, they overlook that their peers are likely examining the social evidence as well (Cialdini, 1984). In our model, players use their predecessors' actions to reduce both the uncertainty about the state and the behaviour of succeeding players while neglecting that others may do so, too.

Definition 2 (Scepticism)**.** Player i is sceptical if (i) she forms an edited history $h_i^S \subseteq h_i$ by removing any action for which there was a majority at the time that action was taken, and (ii) $\hat{\lambda}_{ij} = 1 \geqslant \lambda_j$ for all $j = i+1,\dots,N$.

A *sceptical* player reflects on which of her predecessors could have falsified without taking their conformity motives into account, and removes any action that could represent a falsification. However, she still believes that her successors will be truthful.[8] She is, thus, truthful if $\lambda_i \geqslant \lambda_i(h_i^S,\theta_i) := 1 - [2\big(\theta_i + (1-2\theta_i)\mathbb{E}[\bar{s}_{-i}|h_i^S,\theta_i]\big)]^{-1}$. To build intuition, consider Example 4.1, where $K_j := \sum_{l=1}^{j-1}\mathbb{I}\{s_l = s_j\}$. Player 10 observes history h_{10} and, if sceptical, she removes actions $\{s_4, s_5, s_6, s_8, s_9\}$.[9]

[7] See Section A.3.1 for a derivation of the threshold.

[8] Importantly, neither a sceptical nor a credulous player is myopic: both are imperfectly forward-looking in using a heuristic to simplify the forecasting of successors' actions.

[9] Implicit in our definition of how the sceptical player edits the history is that she posits her predecessors do not edit the history and are, thus, credulous. If instead the sceptical player presumed that a certain predecessor was sceptical and used the same editing rule, there could be cases in which she should not eliminate this predecessor's action although the editing rule tells her to. In Section A.3.2, we provide an alternative model of scepticism that places a sceptical player on a continuum between a fully rational and credulous player.

j	1	2	3	4	5	6	7	8	9
s_j	0	1	0	0	0	0	1	0	0
K_j		0	1	2	3	4	1	5	6
$K_j/(j-1)$		0	1/2	2/3	3/4	4/5	1/6	5/7	6/8
Remove if $\frac{K_j}{j-1} > \frac{1}{2}$	No	No	No	Yes	Yes	Yes	No	Yes	Yes
h_{10}^S	0	1	0				1		

Example 4.1: Editing history h_{10} to h_{10}^S.

For arbitrary histories, we denote the number of removed actions that match, and do not match, player i's type by $A_i^= := \sum_{j=1}^{i-1} \mathbb{I}\{s_j = \theta_i | K_j/(j-1) > 1/2\}$ and $A_i^{\neq} := \sum_{j=1}^{i-1} \mathbb{I}\{s_j \neq \theta_i | K_j/(j-1) > 1/2\}$, respectively. Compared to an interior credulous player $i \in \{2, \ldots, N-1\}$ who underestimates the prevalence of her type based on the history, scepticism decreases the threshold for truth telling: $\lambda_i(M_i, \theta_i) > \lambda_i(h_i^S, \theta_i)$ if $A_i^{\neq} > A_i^=$. Conversely, scepticism increases the cut-off whenever she overestimates the prevalence her type based on h_i.

The concept of salience renders Definition 2 a natural heuristic between credulity and rationality: a player finds it easier to critically reflect on her predecessors' actions as these are salient to her whereas predicting successors' actions remains a difficult task. One reason why past actions may be more salient is due to mental time travel (Tulving, 1985). Past events are typically also viewed more negatively, whilst future events tend to be positively biased (Rasmussen and Berntsen, 2013; Beaty et al., 2019). This negative interpretation of past events may lead a sceptical player to be more critical of the history of play, whilst still believing that her successors will truthfully reveal. Definition 2 also builds on experimental evidence that subjects do make situational attributions for others' behaviour when they can take their role and see the world as they saw it (Taylor and Fiske, 1978). Given knowledge of the history, a sceptical player can reconstruct a predecessor's point of view. A successor's point of view, in contrast, remains inaccessible because a sceptical player suffers an informational deficit relative to the successor. This informational deficit in historical data leads a sceptical player to make trait rather than situational attributions for successors' behaviour (Jones and Nisbett, 1987).

Definition 3 (Rationality)**.** Player i is rational if (i) $\hat{\lambda}_{ij} = \lambda_j$ for all $j \in \mathcal{N}$, and (ii) she knows $\hat{\lambda}_{ji}$ for all $j \in \mathcal{N}$.

A *rational* player takes into account every other player's true conformity motive as well as every other player's presumption about her conformity motive. Hence, in contrast to credulous and sceptical players, a rational player takes into account that a later player's action is determined by both her type and the average action of others'. A rational player is, thus, unable to use the heuristics employed by credulous and sceptical players to simplify (4.1). Instead, the decision to truthfully reveal is equivalent to $\lambda_i \geqslant \lambda_i^R(h_i, \theta_i)$,[10] where

$$\lambda_i^R(h_i, \theta_i) := \frac{\mathbb{E}\big[(\theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, \theta_i\big] - \mathbb{E}\big[(1 - \theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, 1 - \theta_i\big]}{1 + \mathbb{E}\big[(\theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, \theta_i\big] - \mathbb{E}\big[(1 - \theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, 1 - \theta_i\big]}.$$

Dual-process theories render Definition 3 the natural complement to the two preceding heuristics, arguing that there exists two modes of information processing (e.g., Chaiken and Trope, 1999). In addition to a fast, heuristic-driven *System 1* reasoning style, a player may have the time and ability to deploy a slow *System 2* reasoning style.

4.3 Individual Behaviour

To study the effect of boosting in various environments, we begin with a characterisation of individual behaviour. We focus on the conditions under which a player using a heuristic falsifies her type, and how the number of predecessors in agreement with a player's type interact with her heuristic to influence her willingness to falsify. Define $\hat{\pi}(N) := \{\pi | \lambda_i(h_i^S, \theta_i) = 0, (i-1)/2 \leqslant M_i < (i-2+A_i^= - A_i^{\neq})/2\}$ and $\check{\pi}(N) := \{\pi | \lambda_i(h_i^S, \theta_i) = 0, (i-2+A_i^= - A_i^{\neq})/2 < M_i < (i-1)/2\}$.

Lemma 1. *In period $i \in \{1, 2, \ldots, N\}$, if*

C *(i) $M_i \geqslant (i-1)/2$, a credulous player i truthfully reveals her type;*

(ii) $M_i < (i-1)/2$, a sufficiently conformist credulous player i falsifies her type.

S *(i) $M_i \geqslant \max\{(i-1)/2, (i-2+A_i^= - A_i^{\neq})/2\}$, a sceptical player i truthfully reveals her type;*

(ii) $(i-1)/2 \leqslant M_i < (i-2+A_i^= - A_i^{\neq})/2$, a sufficiently conformist sceptical player i falsifies her type in an informative environment $\pi > \min\{\hat{\pi}(N), 1\}$;

(iii) $(i-2+A_i^= - A_i^{\neq})/2 < M_i < (i-1)/2$, a sufficiently conformist sceptical player i falsifies her type in an uninformative environment $\pi < \min\{\check{\pi}(N), 1\}$;

(iv) $M_i \leqslant \min\{(i-2)/2, (i-2+A_i^= - A_i^{\neq})/2\}$, a sufficiently conformist sceptical player i falsifies her type.

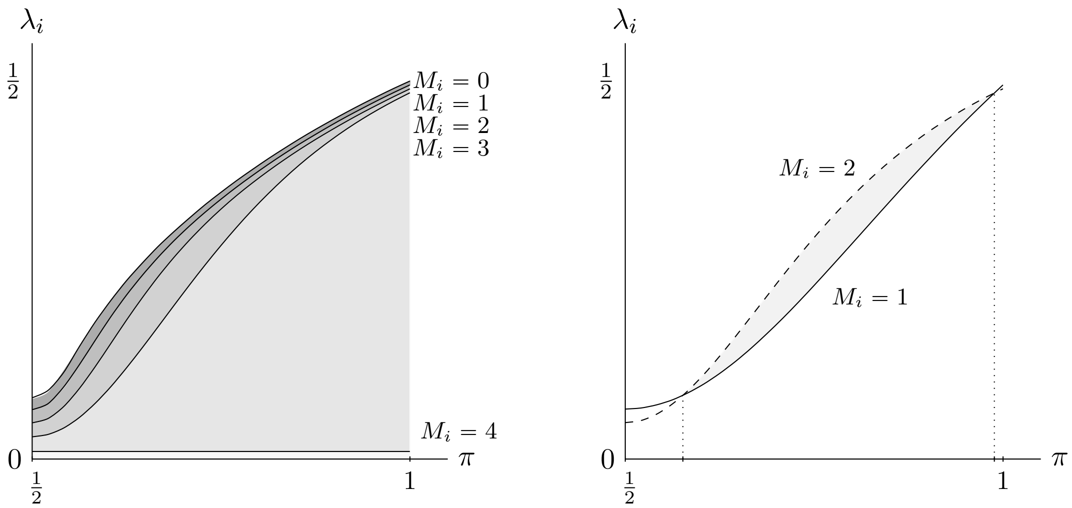
Lemma 1 highlights that the first player, whether credulous or sceptical, always truthfully reveals her type. Similarly, any credulous player $i > 1$ always takes an action that matches her type if it matches the present majority. In other words, a simple majority is

[10] See Section A.3.1 for a derivation of the threshold.

sufficient for a credulous player i to be truthful; a universal consensus is not required. Conversely, when the majority of predecessors disagrees with i's type, $M_i < (i-1)/2$, the social pressure can be strong enough for a credulous player i to falsify: as long as player i thinks herself to be in the minority at period i, she falsifies her type if her weight on conforming is sufficiently high. Our result resonates with Miller and McFarland (1987) in the context of social norms: for widespread conformity to conventions and customs that lack private support, it is only necessary for the majority to think that the majority supports it.

The necessary weight for a credulous player i to falsify is decreasing in the size of the minority. To build intuition, consider an example with $N = 100$. Figure 4.2a shows the cut-off for the tenth player: for all π, she needs to be relatively more conformist to falsify her type when one, two or three predecessors agree with her type rather than none ($M_{10} = 0$). Moreover, when the tenth player can pivot a present majority into a tie by revealing her type ($M_{10} = 4$), her weight on conforming needs to be very large: $s^*_{10}(\theta_{10}) = 1 - \theta_{10}$ if $\lambda_{10} < \lambda_{10}(4, \theta_{10}) = 1/100$. In general, $\lambda_i(M_i, \theta_i)$ approaches $1/N$ as M_i tends to $(i-1)/2$. Therefore, this comparative static highlights that player i is less likely to falsify her type when she belongs to a sizeable current minority.

The same conclusion need not hold for a sceptical player i.[11] A more sizeable minority can increase the cut-off in moderately informative environments: Figure 4.2b provides an example when the first player in the history $h_{10} = \{0, 1, 0, 0, 0, 0, 0, 0, 0\}$ instead announces unity so that the edited history $h^S_{10} = \{0, 1, 0\}$ changes to $h^{S'}_{10} = \{1, 0, 0, 0\}$.



(a) If credulous. (b) If sceptical.

Figure 4.2: The cut-off for player $i = 10$ when $N = 100$.

[11] Importantly, a sceptical player cares about M_i in calculating the expected average action. She uses the edited history only for calculating conditional state probabilities.

4.3.1 Effect of Informativeness

We use Lemma 1 to analyse how individual behaviour depends on the informativeness of a player's type about the state. This allows us to answer our first research question: do more informative environments cause a credulous and a sceptical player to behave differently? If so, a policy maker may be able to use a targeted *scepticism boost* to increase truth telling. Such a boost could be implementable as it does not require the policy maker to know the players' conformity motives.

To boost a credulous player, she could be provided with a decision tree to assist in editing the history (Martignon et al., 2011b,a, Figure 4.3). Alternatively, she could be encouraged to reflect on the history of actions in a short essay. To trigger situational rather than trait attributions provided the action was in the majority at the time, a player could be instructed to consider which of her predecessors she would have expected to behave identically, and which differently, if that player had only cared about fitting in. The essay would help a player to understand whether her presumption $\lambda_{ij} = 1$ and the alternative presumption $\lambda_{ij} = 0$ about predecessor j would be observationally equivalent.

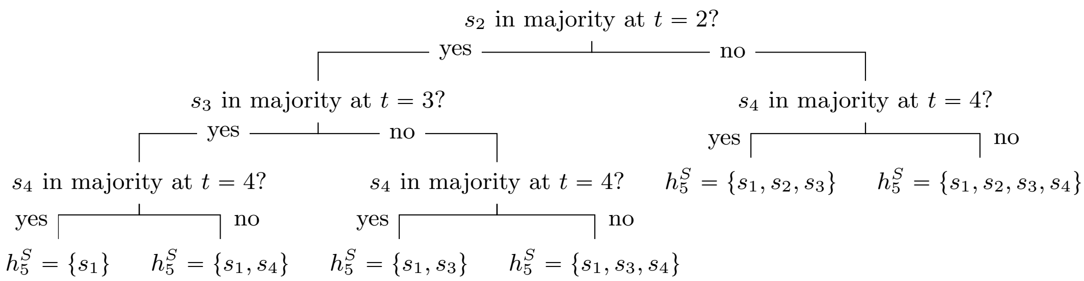


Figure 4.3: Scepticism boost for a credulous player 5 via a decision tree.

The first observation when we compare the behaviour of a credulous player to her sceptical counterpart is that $\lim_{\pi\downarrow 1/2} \lambda_i(M_i, \theta_i) = \lim_{\pi\downarrow 1/2} \lambda_i(h_i^S, \theta_i)$. When players' types are uninformative about the state, they cannot use their predecessors' actions to predict their successors' actions. Therefore, editing the history has no effect on behaviour.

If $M_i \geqslant \max\{(i-1)/2, (i-2+A_i^{=} - A_i^{\neq})/2\}$, player i always chooses the action that matches her type. The behaviour of a credulous and sceptical player coincide for any degree of informativeness and any group size. The credulous player could have under- or overestimated the prevalence of her type based on history h_i. Any overestimation, however, is not severe: the sceptical player still believes the state that matches her type to be weakly more likely based on her edited history and type. If $(i-1)/2 \leqslant M_i < (i-2+A_i^{=} - A_i^{\neq})/2$, the credulous player is always truthful; however, the sceptical player may falsify her type in sufficiently informative environments, given a large enough group size. In particular, for the sceptical player to falsify we require that $\pi > \hat{\pi}(N)$. Given that this threshold on the informativeness is decreasing in N, there exists a minimum group size $\hat{N}$ such that for $N > \hat{N}$, we have $\hat{\pi}(N) < 1$. The minimum group size that solves $\hat{\pi}(\hat{N}) = 1$ is $\hat{N} = 2M_i + 1$.

If $(i-1)/2 > M_i$, a credulous player may choose the action that falsifies her type. Therefore, scepticism may induce this player to change from truth telling to falsification, or vice versa. A change to falsification necessitates an overestimation of the prevalence of her type based on the unedited history; a change to truth telling necessitates an underestimation. If the underestimation is severe, $(i - 2 + A_i^{=} - A_i^{\neq})/2 < M_i < (i-1)/2$, a sufficiently informative environment, given a large enough group size, ensures that the sceptical player is truthful. To guarantee truthful play we require that $\pi > \breve{\pi}(N)$. When the minority of preceding actions match the credulous player's type, $\breve{\pi}(\check{N}) = 1$ is solved by minimum group size $\check{N} = 2(i - 1 - M_i) + 1$.

Taken together, the above observations lead to the following result.

Proposition 1. *Suppose the group is sufficiently large, $N > \hat{N}$. If*

(i) $(i-1)/2 < M_i < (i-2+A_i^{=} - A_i^{\neq})/2$ *for type θ_i, there exists a range of environments,* $\pi \in (\breve{\pi}(N; 1-\theta_i), \hat{\pi}(N; \theta_i)]$, *in which both types of the sceptical player i tell the truth;*

(ii) $\max\{(i-1)/2 + \varepsilon, (i - 2 + A_i^{=} - A_i^{\neq})/2\} \leqslant M_i < (i + A_i^{=} - A_i^{\neq})/2$ *for type θ_i, there exists a range of environments, $\pi \in (\breve{\pi}(N; 1-\theta_i), 1)$, in which both types of the sceptical player i tell the truth.*

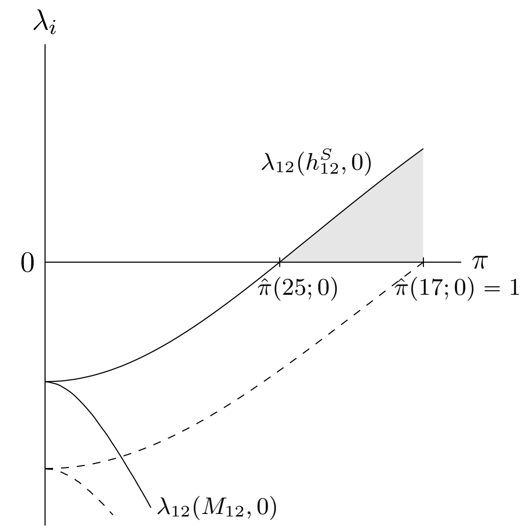
To build intuition about the effect of scepticism on behaviour in Proposition 1(i), consider $h_{12} = \{0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1\}$ with $h_{12}^S = \{0, 1, 1, 1\}$. For the low type of player 12, $\hat{N} = 2 \times 8 + 1$; for the high type, $\check{N} = 2(12 - 1 - 3) + 1$. Therefore, the minimum group size such that type $\theta_{12} = 0$ may falsify for $\pi > \hat{\pi}(N; 0)$ and type $\theta_{12} = 1$ is always truthful for $\pi > \breve{\pi}(N; 1)$ is 17. The informativeness threshold of type $\theta_{12} = 1$ is lower compared to type $\theta_{12} = 0$ when player 12 is part of a group with more then 17 players.

Given this insight, a policy maker can boost a credulous player to become sceptical to increase truth telling.[12] The policy maker does not need to know the player's type or her conformity motive. He only needs to know the informativeness of the environment, which is assumed to be common knowledge. However, the degree of informativeness is crucial for the success of the boost. If the environment is relatively informative, the boost can have unintended consequences if the targeted player turns out to have the low type. She may then falsify if her concern about fitting in is high (Figure 4.4a). Similarly, for less informative

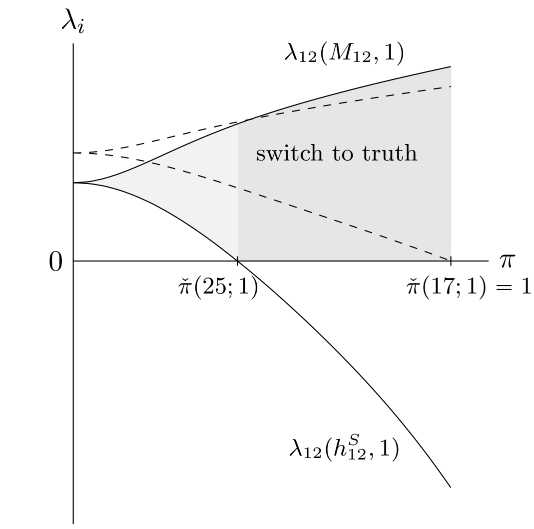
[12]In boardrooms and on committees a policy maker may wish to increase truth telling, as discussed further in Section 4.6.1. We admit, however, that a policy maker does not necessarily wish to increase truth telling universally and, thus, provide a formal social welfare analysis given a utilitarian objective in Section 4.4.2.

environments, the boost may not sway a relatively conformist player with the high type into truth telling (Figure 4.4b). Therefore, for the boost to be effective, the environment must be moderately informative.

Example 4.1 illustrates Proposition 1(ii). Because the edited history ties, the low type of player 10 is always truthful. In this case, the boost can also be effective in highly informative environments.



(a) Sceptical $\theta_{12} = 0$ may falsify for $\pi > \hat{\pi}(25; 0)$.



(b) Sceptical $\theta_{12} = 1$ is always truthful for $\pi > \check{\pi}(25; 1)$.

Figure 4.4: The effect of scepticism on behaviour for $N = 25$ (dashed line for $\hat{N} = \check{N}$).

4.3.2 Comparison to Rational Players

To build towards our main result in Section 4.4, we benchmark our preceding analysis against a rational player who faces other rational players. We focus on individual behaviour in a group of four players for several reasons: (i) given the strategic externality that a rational player understands herself to have on later players' actions, the backward and "forward" induction quickly becomes intractable;[13] (ii) the smallest number of players that allows us to derive interesting insights while keeping a rational player's inference problem tractable is four because only then does there exist an interior player who can have at least one predecessor who did not necessarily truthfully reveal. Given $N = 4$, we compare the behaviour of a credulous player to her rational counterpart to predict whether rationality reduces a player's incentive to falsify.[14]

[13] To prove Lemma 2 for N players, solving backwards, we would have to consider 2^{N-2} histories to calculate the last player's possible cut-offs.

[14] We could also compare the behaviour of a sceptical player to her rational counterpart. As player 2 and 4 do not edit the history, only the analysis of player 3 would differ in Lemma 2(iii) with $\lambda_3(h_{30}^S, \theta_3) =$

Lemma 2. *Under rationality:*

(i) player 1's behaviour is unchanged.

(ii) player 2's cut-off is weakly higher when $M_2 = 0$.

(iii) player 3's cut-off

(a) is weakly higher when $s_2^*(\theta_2) = \theta_2$ *and* $M_3 = 0$.

(b) can be higher or lower when $s_2^*(\theta_2) = \theta_1$ *and* $M_3 = 0$.

(iv) player 4's cut-off is unchanged.

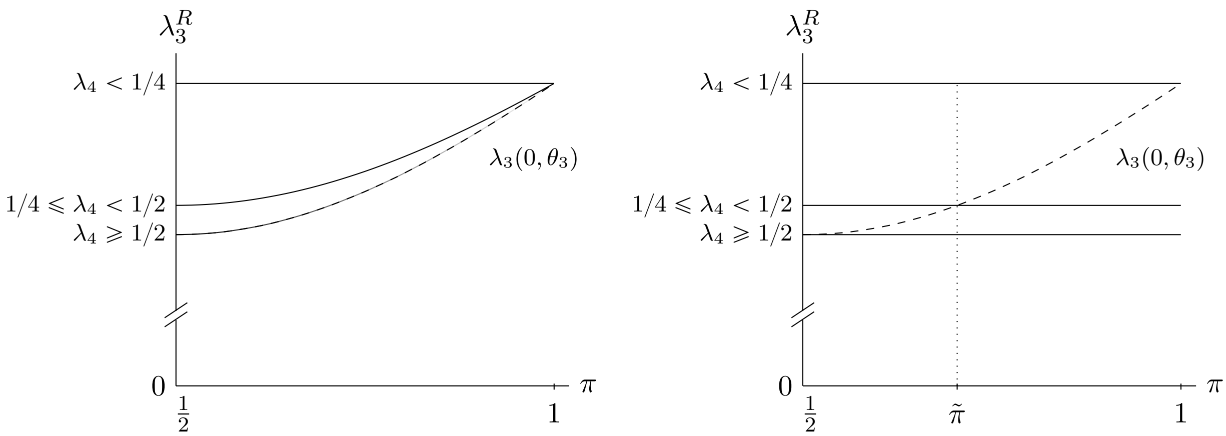
A rational player 1 also tells the truth due to evidence in favour of the state that matches her type and the symmetric environment. When every successor is highly conformist, player 1 does not face a trade-off between matching her type and matching the average action. When her successors are instead relatively nonconformist, player 1 maximises the probability of matching her action to her successors' types by revealing truthfully. Simultaneously, she avoids disutility from falsifying her type. When player 2 is always truthful, $s_1(\theta_1) = \theta_1$ is optimal because it maximises the probability of matching player 2's type, thereby maximising the probability of a history that induces later conformist players to agree. When player 2 always agrees with player 1, $s_1(\theta_1) = \theta_1$ increases the probability of her action matching some later nonconformist player's type while avoiding the disutility from falsifying.

Given this insight, a rational player 2 is in a special position: contrary to player 3, her predecessor always tells the truth. Therefore, the use of a heuristic does not alter the conditional state probabilities in the calculation of player 2's expected utility. Moreover, when player 2 disagrees with player 1, her action does not have an effect on her direct successor. Whether later successors falsify depends on the evolution of the history. Given that each state is equally likely and the symmetric environment, player 2 expects $\mathbb{E}[(s_2 - \bar{s}_{-2})^2 | s_1 \neq s_2 = \theta_2]$ to take on the same value as if all successors were truthful. In contrast, when a rational player 2 agrees with player 1, she realises that someone after her with the same type might also falsify. This reduces the expected distance between $\bar{s}_{-2}$ and action $s_2 = s_1 \neq \theta_2$. As a result, the rational player 2's cut-off is higher as long as her successors do not always tell the truth. Therefore, the use of a heuristic does not exacerbate falsification for player 2: the rational player 2 can be less conformist and still falsify.

A rational player 3 takes into account when she can infer player 2's type from her action. If she considers that player 2 is truthful, she is in the same position as player 2 with

$2/5 < \lambda(0, \theta_3)$: the rational cut-off is strictly higher than the sceptical if $s_2^*(\theta_2) = \theta_2$, and weakly higher if $s_2^*(\theta_2) = \theta_1$.

a higher rational cut-off as long as her successor does not always tell the truth (Figure 4.5a). If she considers that player 2 always agrees with player 1, she has less evidence in favour of the state that contradicts her type. If player 4 always agrees with the majority ($\lambda_4 < 1/4$), credulity reduces the cut-off; if player 4 always tells the truth ($\lambda_4 \geqslant 1/2$), credulity increases the cut-off (Figure 4.5b). Hence, credulity exacerbates falsification if there is no strategic externality on successors' actions so that this heuristic amounts to inferential naivety (Eyster and Rabin, 2010) - a player positing that her predecessors' actions only reflect their types. Put differently, the presence of a strategic externality on successors' actions in addition to a potential informational externality arising from incorrectly inferring the state voids the conclusion that the use of heuristics always exacerbates falsification (Bohren, 2016).



(a) If $s_2^*(\theta_2) = \theta_2$.

(b) If $s_2^*(\theta_2) = \theta_1$.

Figure 4.5: The cut-off for player 3 under rationality when $N = 4$.

The change to rationality does not affect player 4's cut-off. Her ability to observe the actions of all other players means that, whether she presumes all other players to truthfully reveal or not, her incentive to falsify remains unchanged. The observation that the last player's cut-off is not affected by the use of a heuristic generalises to any group size as the last player does not need to calculate conditional state probabilities, and cannot mispredict successors' actions.

Lemma 2(ii)-(iii) suggest a general pattern for interior players. If a credulous player $i \in \{2, \dots, N-1\}$ correctly interprets her predecessors' actions but neglects the social pressure of a sufficiently uniform history on her successors' actions, then the rational cut-off is higher than the credulous cut-off. Intuitively, underestimating the social pressure on successors leads the credulous player i to feel less social pressure herself. If instead a credulous player i neglects the social pressure in determining her predecessors' actions, but correctly infers that neither she nor her predecessors have a strategic externality on her successors' actions, then the rational cut-off is lower. Intuitively, overestimating the prevalence of type $1-\theta_i$ based on the history leads the credulous player i to feel more social pressure.

Said pattern complements Kahan's (2013) finding that, relative to a heuristic reasoning style, an effortful System 2 reasoning style tends to magnify identity-protective cognition - the fitting of one's action to the group's expected average action. In particular, we show that when players move sequentially, the validity of this conclusion depends on whether the heuristic reasoning style affects a player's understanding of the history or forecasting of the future. Moreover, the pattern corroborates Kuran and Sunstein's (1999) prediction that an (availability) heuristic interacts with the boundaries of acceptable nonconformism.

Lemma 2 provides insights for policy. A *rationality boost* only affects the behaviour of interior players. Moreover, this boost can have heterogeneous effects: a boost that induces player 2 to always agree with player 1 instead of truthfully revealing her type does not necessarily have the same effect on later players. Given $1/4 \leqslant \lambda_4 < 1/2$ and $\pi > \tilde{\pi}$, if $\lambda_3^R(h_{30}, \theta_3) \leqslant \lambda_3 < \lambda_3(0, \theta_3)$ and $\lambda_2(0, \theta_2) \leqslant \lambda_2 < \lambda_2^R(h_{20}, \theta_2)$, where $h_{30} = \{1-\theta_2, 1-\theta_2\}$ and $h_{20} = \{1-\theta_2\}$, respectively, player 3 switches from falsifying to truthtelling but player 2 switches from truthtelling to falsifying. Therefore, the elimination of heuristic decision-making does not guarantee that more players truthfully reveal their types. An intervention that lowers the pressure to conform in the presence of heuristic decision-making (Section 4.6.1) may be more effective. If nonetheless chosen by a policy maker, a rationality boost could, in practice, be implemented by disseminating infographics (e.g., Spiegelhalter et al., 2011); for example, a unit interval displaying where a player's conformity motive lies relative to each of her peers'.

4.4 Aggregate Behaviour

We use our results on individual behaviour from Section 4.3 to derive aggregate behaviour in a group of four players. Given that (*i*) the group is small, which could model close friends, and (*ii*) individuals tend to associate with others who are similar to themselves so that they share behavioural characteristics (McPherson et al., 2001), we consider players with homogeneous conformity motives: $\lambda := \lambda_i = \lambda_j$ for all $i \neq j \in \mathcal{N}$. This assumption also allows us to focus entirely on the role of players' presumptions in order to answer our main research question: does a more complex heuristic make uniform behaviour less likely? To compare outcomes across credulous, sceptical and rational groups, we first introduce the concept of locking in an action from period k.[15]

Definition 4 (Lock-in)**.** An action is locked in from period k if, and only if,

(i) $s_j^*(\theta_j) = s_k^*(\theta_k)$ for all $j \in \{k+1, \ldots, N\}$, and

(ii) $K_k > (k-1)/2$.

Lock-in necessitates players k to N to take the same action. This action has to have a majority at period k: condition (ii) delays lock-in to a period k in which s_k could be taken by a credulous player of either type. A direct implication of condition (ii) is that lock-in cannot occur in period 1 because the first player always tells the truth. Further, condition (ii) implies that the edited history does not change under lock-in: sceptical players k to N face the same edited history. Consequently, the increase in social pressure from moving one period later after that an action has been locked in is lower for sceptical than for credulous players. The effect of lock-in under credulity formalises Kuran and Sunstein's (1999) argument that the availability heuristic interacts with social mechanisms to generate a cascade that (i) "progressively frees public discourse of voices out of tune with the evolving chorus", and (ii) "facilitates [players'] participation in the evolving consensus" by "making their private preferences seem increasingly unusual" (p.731).

Theorem 1. *The mapping between types and actions, $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (s_1, s_2, s_3, s_4)$, is:*[16]

(i) $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *if*
$$\begin{cases} \lambda \in \big[0, \lambda_2(0, \theta_2)\big) & \textit{when credulous or sceptical,} \\ \lambda \in \big[0, \lambda_2^R(h_{20}, \theta_2)\big) & \textit{when rational,} \end{cases}$$
where $\lambda_2^R(h_{20}, \theta_2) > \lambda_2(0, \theta_2)$.

(ii) $(\theta, \theta, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $(\theta, 1-\theta, \theta_3, \theta_4) \mapsto (\theta, 1-\theta, \theta_3, \theta_4)$ *if*
$$\begin{cases} \lambda \in \big[\lambda_2(0, \theta_2), \lambda_3(0, \theta_3)\big) & \textit{when credulous,} \\ \lambda \in \big[\lambda_2(0, \theta_2), \lambda_3(h_{30}^S, \theta_3)\big) & \textit{when sceptical,} \\ \lambda \in \big[\lambda_2^R(h_{20}, \theta_2), \lambda_3^R(h_{30}, \theta_3)\big) & \textit{when rational,} \end{cases}$$
where $\lambda_3^R(h_{30}, \theta_3) > \lambda_3(0, \theta_3) > \lambda_3(h_{30}^S, \theta_3)$.

(iii) $(\theta, \theta, \theta, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $\{(\theta, \theta, 1-\theta, \theta_4), (\theta, 1-\theta, \theta, \theta_4), (\theta, 1-\theta, 1-\theta, \theta_4)\} \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if*
$$\begin{cases} \lambda \in \big[\lambda_3(0, \theta_3), \lambda_4(0, \theta_4)\big) & \textit{when credulous,} \\ \lambda \in \big[\lambda_3(h_{30}^S, \theta_3), \lambda_4(0, \theta_4)\big) & \textit{when sceptical,} \\ \lambda \in \big[\lambda_3^R(h_{30}, \theta_3), \lambda_4(0, \theta_4)\big) & \textit{when rational.} \end{cases}$$

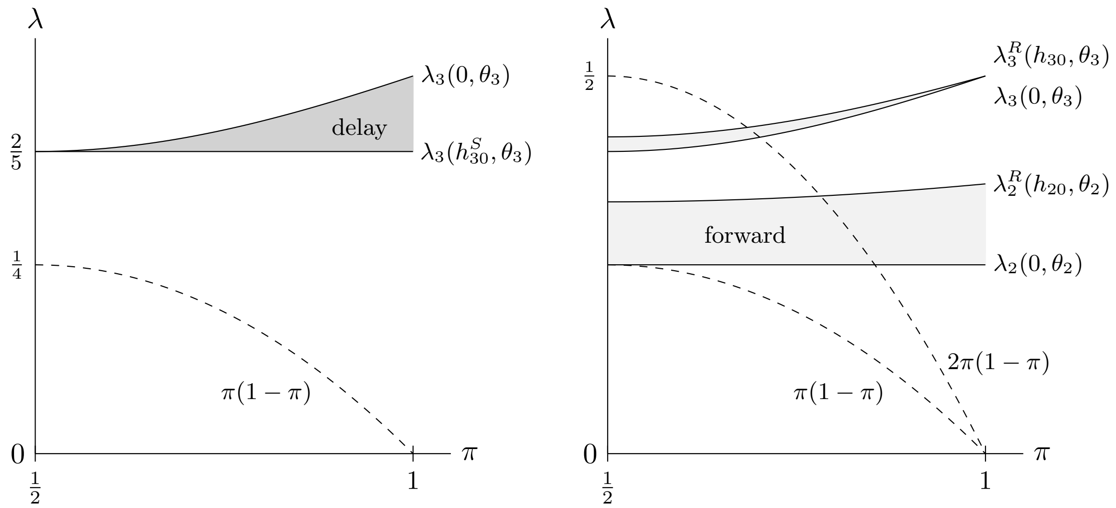
[15] We show in Section A.3.3 that the insights from Section 4.4 are robust to heterogeneity in the level of rationality within the group.

[16] To simplify the exposition, we set $\theta_1 = \theta$. For example, $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ means that (i) player 1 truthfully reveals, $s_1^*(\theta) = \theta$, and (ii) players 2, 3 and 4 agree with player 1, no matter what their types are: $s_j^*(\theta_j) = \theta$ for $j = 2, 3, 4$.

(iv) $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if* $\lambda \in \left[\lambda_4(0, \theta_4), 1\right]$ *when credulous, sceptical or rational.*

Theorem 1(iii) highlights that a change from credulity to scepticism may delay the possibility of lock-in by one period to $k = 4$. Furthermore, for $\lambda_3(h_{30}^S, \theta_3) \leqslant \lambda < \lambda_3(0, \theta_3)$, the probability of observing uniform behaviour falls by $\pi(1-\pi)$. While a more informative environment increases the range of conformity motives for which lock-in can be delayed, the effect on the probability of observing uniform behaviour is smaller (Figure 4.6a).

Theorem 1(i) highlights that a change from credulity to rationality may bring forward the possibility of lock-in by one period to $k = 2$. Furthermore, for $\lambda_2(0, \theta_2) \leqslant \lambda < \lambda_2^R(h_{20}, \theta_2)$, the probability of observing uniform behaviour increases by $2\pi(1-\pi)$. Theorem 1(ii) highlights that a change to rationality can also bring forward the possibility of lock-in to $k = 3$. For $\lambda_3(0, \theta_3) \leqslant \lambda < \lambda_3^R(h_{30}, \theta_3)$, the probability of observing uniform behaviour increases by $\pi(1-\pi)$. In this case, a less informative environment increases both the range of conformity motives for which lock-in can be brought forward and the effect on the probability of observing uniform behaviour (Figure 4.6b).



(a) Scepticism boost.

(b) Rationality boost.

Figure 4.6: Effect of boosts in parameter space (π, λ).

The preceding discussion reveals that a scepticism boost in the aggregate only ever has an effect in a moderately conformist group. This result contrasts with the effectiveness of a targeted scepticism boost in Section 4.3.1, where a policy maker only needed to know the informativeness of the environment. For effectiveness at the aggregate level, he also needs to know the common conformity motive. A rationality boost has a close-to-maximal effect in a relatively or moderately conformist group when the environment is uninformative.

The comparison of the boosts highlights that scepticism is not sufficient to incentivise players to adopt a positive social norm, such as pro-climate behaviour: to lock in the action corresponding to the state of the world, a boost to rationality is necessary. In addition, policy maker needs to know, or be able to influence, what type of player moves first.

We can use Theorem 1 to derive the ex-ante probability of uniform behaviour. This summary statistic allows us to analyse the expected effect of a boost to scepticism or rationality prior to the interaction. To facilitate the result, we assume that the common conformity motive is drawn from a continuous distribution F with full support.

Corollary 1. *For any F, a scepticism (rationality) boost reduces (increases) the probability of uniform behaviour.*

Given that F is strictly increasing on the unit interval, uniform behaviour is always more (less) likely in a credulous than a sceptical (rational) group. Hence, a policy maker can use a scepticism boost to reduce the ex-ante probability of uniform behaviour and a rationality boost to increase it. Depending on the distribution of conformity motives, the effect of a rationality boost can be larger or smaller than the effect of a scepticism boost. In the special case that all conformity motives are equally likely prior to the interaction, $\lambda \sim \mathcal{U}[0, 1]$, the change in the ex-ante probability of uniform behaviour from a rationality boost is larger.

4.4.1 Pluralistic Ignorance

If players are credulous, lock-in can generate a false consensus effect: when $k = 2$ (Theorem 1(i)), a credulous player who chooses an action that matches her type concludes that all other players share her type. She mistakenly concludes her type to be the consensus. Every credulous player who falsifies her type, in contrast, concludes that she is the only deviant. While she falsifies her type, she concludes that no one else does, so that the heuristic triggers "illusory feelings of deviance" (Miller and Prentice, 1994).

Credulous players who falsify their types suffer from *pluralistic ignorance*: credulity leads them to "infer that the identical actions of the self and others reflect different internal states" (Miller and McFarland, 1987, p.298).[17] To formalise pluralistic ignorance, we adopt a broad definition that is based on an underestimation of the prevalence of one's type at the time of one's choice.

Definition 5 (Pluralistic ignorance)**.** Player i's strategy to falsify her type is pluralistically ignorant if $\lambda_i(M_i, \theta_i) > \lambda_i \geqslant \lambda_i(h_i^S, \theta_i)$.

[17] Pluralistic ignorance has also been shown to arise when players adhere to a social norm based on the false presumption that others will enforce it (Centola et al., 2005).

A player's strategy is pluralistically ignorant if she only falsifies her type because she takes the history of play at face value. In our model, pluralistic ignorance is, thus, essentially a form of social availability error (Kuran and Sunstein, 1999), triggered by the interaction of an (availability) heuristic and the social mechanism of conformity. A number of implications follow from Definition 5. First, a credulous player has to be in the minority when she makes her choice. She would never choose to falsify her type when $M_i \geqslant (i-1)/2$. Second, a credulous player must underestimate the prevalence of her type.[18]

4.4.2 Social Welfare

So far, we discussed informally the possibility that a policy maker wishes to (i) increase truth telling, or (ii) incentivise players to adopt a positive social norm. We now use Theorem 1 to offer a more formal welfare analysis in the group of four players with homogeneous conformity motives when the policy maker's objective is to maximise the sum of players' utilities.

This utilitarian objective may be appropriate if the policy maker has no agenda of his own, say, to elicit the true distribution of types in the group to preserve diversity in public, or remain capable of acting given the principle of unanimity.[19] Hence, when ex-ante neither extreme of universal truth telling nor uniform behaviour with certainty is preferable to the other possible outcomes, a policy maker may wish to treat all players equally by defining $W = \sum_{i \in \mathcal{N}} \mathbb{E}[U(s_i;\theta_i,\lambda_i)]$.

Corollary 2. *A utilitarian policy maker prefers*

(i) $(\theta,\theta_2,\theta_3,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ *if* $\lambda \in [0,\lambda')$,

(ii) $(\theta,\theta,\theta_3,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ *and* $(\theta,1-\theta,\theta_3,\theta_4) \mapsto (\theta,1-\theta,\theta_3,\theta_4)$ *if* $\lambda \in (\lambda',\lambda'')$,

(iii) $(\theta,\theta,\theta,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ *and* $\{(\theta,\theta,1-\theta,\theta_4),(\theta,1-\theta,\theta,\theta_4),(\theta,1-\theta,1-\theta,\theta_4)\} \mapsto (\theta,\theta_2,\theta_3,\theta_4)$ *if* $\lambda \in (\lambda'',\lambda''')$,

(iv) $(\theta,\theta_2,\theta_3,\theta_4) \mapsto (\theta,\theta_2,\theta_3,\theta_4)$ *if* $\lambda \in (\lambda''',1]$,

[18] Our insights echo Bicchieri (2005, p.187-188):

> pluralistic ignorance occurs when individuals engage in social comparison and other people's behavior is observable, but there is no transparent communication, in that public stances may differ from private attitudes and preferences. Individuals, however, will assume that the observed behavior is consistent with the actors' underlying attitudes and preferences. Hence, when the observed behavior points to the existence of a shared norm, people in the grip of pluralistic ignorance will tend to conclude that the norm is widely endorsed and feel compelled to conform.

[19] The Council of the EU (2020), for example, has to vote unanimously on a number of matters which the member states consider to be sensitive.

where $\lambda' := [6 + 4\pi(1-\pi)]/[15 + 4\pi(1-\pi)] > \lambda_2^R(h_{20}, \theta_2)$*,* $\lambda'' := [12 + 8\pi(1-\pi)]/[21 + 26\pi(1-\pi)] > \lambda_3^R(h_{30}, \theta_3)$ *and* $\lambda''' := 4/7 > \lambda_4(0, \theta_4)$*.*

Given that $\lambda' > \lambda_2^R(h_{20}, \theta_2)$, a utilitarian policy maker prefers uniform behaviour with certainty, that is, the first mapping in Theorem 1, more often than it obtains in the interaction - whether players are credulous, sceptical or rational. This observation also implies that a utilitarian policy maker prefers the four players to be rational for $\lambda \in [\lambda_2(0, \theta_2), \lambda_2^R(h_{20}, \theta_2))$. Hence, for this range of conformity motives, a utilitarian policy maker achieves the welfare-maximising mapping through a rationality boost.

Similarly, for $\lambda \in [\lambda_3(0, \theta_3), \lambda_3^R(h_{20}, \theta_3))$, a utilitarian policy maker prefers the four players to be rational to bring forward the possibility of lock-in to period 3 and increase the probability of uniform behaviour. For this range of conformity motives, a rationality boost improves the outcome of the interaction from the policy maker's perspective but does not necessarily achieve the welfare-maximising mapping. For sufficiently uninformative environments, the policy maker ideally would want to forward lock-in even further to period 2. For sufficiently informative environments, a rationality boost does maximise social welfare. As an implication of this observation, a scepticism boost cannot be optimal from a utilitarian policy maker's perspective.

4.5 Complete Information

In our analysis so far, players faced uncertainty about the state. An important question for policy is, thus, whether incomplete information is necessary for heuristic decision-making to have its full effect on individual and aggregate behaviour. If complete information about the state is sufficient for preference falsification and lock-in of a certain action, a policy maker may be able to adjust the information structure rather than altering players' presumptions.[20] Therefore, we now revisit our results from Section 4.3 and Section 4.4 under complete information.

When the state is known, it ceases to matter whether a player's predecessors are truthful as their actions are not used for predicting successors' actions. A credulous or sceptical player may only underpredict how uniform her successors' play will be given the history and her action. As editing the history becomes unnecessary, the complete information cut-offs of credulous and sceptical players coincide. Thus, the following characterisation applies

[20] Common knowledge of the state implies independence of players' types but the probability of the types need not be one-half. Alternatively, we could study an uninformative environment (i.e., $\pi \to 1/2$). However, this approach is less flexible as one would approach independence with a uniform probability of the two types.

to both a credulous and sceptical player i. Define $\underline{\pi} := [N-1-2M_i]/[2(N-i)]$ and $\bar{\pi} := [N+1-2(i-M_i)]/[2(N-i)]$.

Lemma 1 (Complete information)**.** *In period $i \in \{1,\dots,N-1\}$, if $\theta_i = \omega$ and*

(i) $M_i \geqslant (i-1)/2$, *player i truthfully reveals her type;*

(ii) $(2i-N-1)/2 < M_i < (i-1)/2$, *a sufficiently conformist player i falsifies her type in an uninformative environment $\pi < \underline{\pi}$;*

(iii) $M_i \leqslant (2i-N-1)/2$, *a sufficiently conformist player i falsifies her type;*

In period $i \in \{1,\dots,N-1\}$, if $\theta_i \neq \omega$ and

(iv) $M_i \geqslant (N-1)/2$, *player i truthfully reveals her type;*

(v) $(i-1)/2 < M_i < (N-1)/2$, *a sufficiently conformist player i falsifies her type in an informative environment $\pi > \bar{\pi}$;*

(vi) $M_i \leqslant (i-1)/2$, *a sufficiently conformist player i falsifies her type.*

Lemma 1 (Complete information) highlights that the first player may falsify her type under complete information. If she knows her type to differ from the state and neglects her successors' conformity motive, she expects to match most successors' actions through falsification. Hence, knowledge of the state prevents a sufficiently conformist first player from locking in the action that does not match the state from period 2.

Compared to when the state is unknown, complete information introduces more factors in a credulous player's decision rule. Cases (*i*) and (*vi*) are similar to Lemma 1**C**: when at least half of the preceding players have agreed with player i's type that matches the state, she truthfully reveals; conversely, when at least half have disagreed with the type that does not match the state, a sufficiently conformist player falsifies. As long as at least half of *all* other players have agreed with player i's type that does not match the state (case (*iv*)), even a conformist player is guaranteed to truthfully reveal. Moreover, as long as player i is in a relatively small minority (case (*iii*)), falsification is possible in any environment even when player i knows her type to match the state.

When the number M_i of agreeing predecessors is in an intermediate range, falsification is possible in certain environments. When player i's type does match the state (case (*ii*)), she needs to be sufficiently conformist and the environment needs to be sufficiently uninformative for the probability of many successors announcing θ_i to be perceived to be low. When player i's type does not match the state (case (*v*)), she needs to be sufficiently conformist and the environment needs to be sufficiently informative for the probability of many successors announcing θ_i to be perceived to be low.

Player N's decision whether to falsify is independent of her type matching the state. She observes the full history of play, and does not have to predict her successor's based on the realisation of the state. Therefore, her behaviour reduces to that in Lemma 1.[21]

We can now compare the cut-off of a credulous player $i \in \{1, \dots, N-1\}$ under complete and incomplete information: if her type matches the state, the complete information cut-off is lower; if it does not match, the cut-off is higher. To show this, fix $\theta_i = 1$ without loss of generality. The incomplete information cut-off is $\lambda_i(M_i, 1)$ with $\mathbb{E}[\bar{s}_{-i}|M_i, 1] = M_i/(N-1) + q(N-i)/(N-1)$, where $q := \Pr(\omega = 1|M_i, 1)\pi + \Pr(\omega = 0|M_i, 1)(1-\pi)$. Under complete information, q is replaced by π if the player's type matches the state, and by $1-\pi$ if it does not. Given that $1-\pi < q < \pi$, we have $\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 0] < \mathbb{E}[\bar{s}_{-i}|M_i, 1] < \mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 1]$ which implies $\lambda_i(M_i, 1; \omega = 0) > \lambda_i(M_i, 1) > \lambda_i(M_i, 1; \omega = 1)$.

Further, if the number of predecessors agreeing with a player remains constant, $M_i = M_{i+1}$, the distance between the complete and incomplete information cut-offs is smaller when player i moves in period $i+1$. Intuitively, the state has less importance in the expected average action: there are fewer actions to predict, and only for those actions does knowledge of the state matter. For example, when $\theta_i = \omega = 1$, the expected average action falls by $\pi/(N-1)$ under complete information, but only by $q/(N-1)$ under incomplete information from period i to period $i+1$. Together with the concavity of the player's cut-off in the expected average action and the expected average action being higher under complete information, the conclusion follows.[22]

Given the insights for player 1 to $N-1$, a policy maker who prefers (i) truth telling if a player's type matches the state, and (ii) falsification if it contradicts the state, can make common knowledge which type is more likely held by the majority. Given heuristic decision-making, this intervention will be more effective in sufficiently informative environments.

Under incomplete information, Definition 1 had two effects: an informational externality whenever a credulous player incorrectly inferred the state, and the neglect of a strategic externality on successors' actions. Due to the interaction of the two effects, a rational player's cut-off in Lemma 2 could be higher or lower.

Lemma 2 (Complete information). *Let $i = N-1$ and ω be known. Compared to her credulous or sceptical counterpart, the rational player's cut-off is*

(i) weakly greater if $M_i < (i-1)/2$;

[21] In particular, when $i = N$, cases (i) and (iv) become equivalent, as do cases (iii) and (vi). Moreover, the intervals in cases (ii) and (v) no longer exist.

[22] The same conclusion follows when $\theta_i = 1 \neq \omega$ as $\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 0]$ is lower and falls by less.

(ii) weakly lower if $M_i > (i-1)/2$;

(iii) weakly lower (greater) if her type does not match (matches) the state and $M_i = (i-1)/2$.

Under complete information, the first effect is absent. Rationality only corrects for strategic externalities. Hence, Lemma 2 (Complete information) suggests that a rational player may act differently because she takes into account that social pressure can alter later players' actions. She considers that the social pressure on her successors increases (decreases) when she agrees (disagrees) with most of her predecessors. When her type is in the majority (minority) at h_i, this understanding lowers (increases) her cut-off to truthfully reveal her type. The absence of the first effect, therefore, leads to a clear-cut adjustment following a rationality boost, depending on whether a player's type is in the majority at the time of her decision. In contrast to incomplete information, a policy maker can, thus, unambiguously predict the effects of a rationality boost on individual players' behaviour.

We use our preliminary results to provide new insights on aggregate behaviour under complete information and homogeneous conformity motives.

Theorem 1 (Complete information)**.** *(i) The mapping between types and actions is $(\theta, \theta_2, \ldots, \theta_N) \to (\omega, \omega, \ldots, \omega)$ if $\lambda \in \big[0, (2\pi-1)/(2\pi)\big)$ in a group of credulous or sceptical players.*

(ii) The mapping $(\theta, \theta_2, \ldots, \theta_N) \to (\theta, \theta, \ldots, \theta)$ does not exist if $\pi \geqslant (N-1)/(2N-3)$ in a group of credulous or sceptical players.

First, disseminating information about the state of the world can be an effective alternative to an aggregate rationality boost (Figure 4.6b) to incentivise the adoption of a positive social norm: to ensure lock-in of the action corresponding to the state, a policy maker needs to know that the homogeneous group is sufficiently conformist. In highly informative environments, it suffices to know that the group cares relatively more about conforming than matching one's type.

Second, under incomplete information, a highly conformist group always followed the first player (Theorem 1(i)). Under complete information, in contrast, a sufficiently informative environment prevents lock-in of the action corresponding the the first player's type from period 2. In a four-player example, this informativeness threshold is $\pi = 3/5$. Further, the threshold is decreasing in the number of players so that lock-in of θ from period 2 under complete information becomes essentially impossible in very large homogeneous groups.

4.6 Concluding Remarks

In this paper, we analysed heuristic decision-making. To understand its effect on individual behaviour, we first derived the conditions under which a credulous, sceptical or rational player falsifies her type. Whilst a credulous player is less likely to falsify when she belongs to a sizeable current minority, the same conclusion does not need to hold for a sceptical player. We showed that, by changing the informativeness of the environment, the behaviour of credulous and sceptical players may diverge: credulous players boosted to scepticism may switch from conforming to revealing. Moreover, compared to rationality, credulity does not necessarily exacerbate falsification. Our results suggest that correctly interpreting predecessors' actions but neglecting the social pressure on successors leads a credulous player to feel less social pressure; conversely, neglecting the social pressure on her predecessors but correctly predicting successors' behaviour leads her to feel more.

To understand the effect on aggregate behaviour and answer our main research question, we characterised the mapping between types and actions when four players have homogeneous conformity motives. We showed that a boost from credulity to scepticism can delay the possibility of an action being locked in, whilst a boost to rationality can bring this possibility forward. Scepticism can also reduce the probability of uniform behaviour, whilst rationality can increase it. Moreover, the relative effect of rationality on the probability of uniform behaviour is greater than that of scepticism when all conformity motives are ex-ante equally likely.

Our final analysis examined the implications of making the state of the world common knowledge. Under complete information, credulous and sceptical players' behaviour coincides because predecessors' actions are not used to predict successors' actions. Unlike under incomplete information, the first player may falsify when the state is known. If her type differs from the state, the use of a heuristic may lead her to falsify to match most successors' actions. Generically, if a player's type matches the state, the degree of conformity required for her to falsify is higher; if it does not match, it is lower. Moreover, under complete information, a sufficiently informative environment prevents an action from being locked in from the second period.

Throughout the paper, we considered a setting where players are aware of their type. However, literature in behavioural economics has shown that due to, inter alia, dynamic inconsistency (Laibson, 1997), or context dependency and framing effects (Tversky and Kahneman, 1981), players may not always know their true type. As March (1978) pointed out, "[l]imitations of memory organization and retrieval and of information capacity affect information processing about preferences just as they affect information processing about consequences". Thus, the inclusion of uncertainty about one's type is a promising direction for future research to fully understand the effects and potential interdependencies of these two types of cognitive limitations (i.e., attribution bias and limited recall).

To highlight the implications of our results and the welfare impact of heuristic decision-making, we conclude with two applications that vary the nature of players' actions and types. In some situations, a policy maker will want to boost credulous players to scepticism or reduce conformity to limit pluralistic ignorance, thereby improving collective decision-making. In other situations, rationality, an informative environment, or even common knowledge of the state will be optimal for positive social change: locking in an action early on may be desirable to maximise social welfare.

4.6.1 Avoiding Misinformation

Our model sheds light on organisational decision-making: in boardrooms and on committees, a player's action is naturally interpreted as her public opinion. She also holds a fixed private opinion on the debated issue. In these situations, it is conventional wisdom that, if the chair opens the meeting by declaring what she expects its outcome to be, the attendees are almost guaranteed to agree with her, whether or not they believe the chair's plan to be sensible (Hill, 2018). The attendees are victims of *groupthink*: their "strivings for unanimity override their motivation to realistically appraise alternative courses of action" (Janis, 1972, p.9).

Lemma 1**C** captures three symptoms of groupthink: self-censorship takes the form of falsification, the illusion of unanimity is a false consensus, and pressure on dissenters is a form of social pressure. Given that groupthink is commonly associated with inferior decision-making, illustrated by the Challenger disaster (Janis, 1991), a policy maker may wish to generate a situation in which players feel less pressure to conform.

One option discussed in Section 4.3.1 and Section 4.4 is to boost credulous players to reflect on the history of play without knowing their predecessors' conformity motives. Alternatively, a policy maker can aim to reduce credulous players' weights on conforming. Lemma 1**C** gives the target weight for arbitrary histories. To induce attendees with a view differing from the chair to speak their mind, they need to care less about being judged. Their desire to conform may be diminished through secret ballots, blind vetting, or diversity (Kets and Sandroni, 2021). More generally, small-group interventions include safe spaces and confidential conversations with counsellors. In large groups, indirect survey techniques may help diminish biases from social desirability such as underreporting (Rosenfeld et al., 2016). The surveyor may, for example, guarantee a respondent anonymity.

4.6.2 Social Engineering

A player's decisions to falsify her type creates inefficiencies when an inferior status-quo is locked in. The very same mechanism, however, can enhance welfare (Kuran, 1995) by creating cohesion, silencing minor disagreements, or levering social change; for example, when individual positive behaviour is costly and requires widespread adoption to pay off for society at large. Such *social engineering* can incentivise a re-thinking on important issues like climate change, tax compliance and political correctness, which can lock in a welfare-enhancing equilibrium.

One example is a shared commitment to pro-climate behaviour. As noted by Mildenberger and Tingley (2019), second-order beliefs structure people's climate-related behaviours: what one believes everyone else believes determines whether one acts green. They show that people underestimate both pro-climate beliefs and the support for pro-climate policies, and that support increases after people update their second-order beliefs. Here, we interpret public support for pro-climate policies as a player's action, and second-order beliefs as players' presumptions about other players' types.

In our model, the underestimation can be indicative of little public information regarding the distribution of types: a low informativeness of each player's type about the state. In this case, disseminating information about the distribution of types is required. The intervention reduces the uncertainty about the majority type, and makes collective behaviour change more likely. From the cut-offs derived in Lemma 1**C**, it follows that credulous players in a large group only need to care relatively more about conforming than matching their types to alter their behaviour in highly informative environments. Consequently, a policy maker can either induce a norm of pro-climate behaviour or avoid a "welfare-damaging equilibrium condition where individuals fail to coordinate even though all desire some form of political action because they believe that others do not share their willingness to act" (Mildenberger and Tingley, 2019, p.1285).

Empirical evidence on the intervention's effectiveness is provided by Geiger and Swim (2016): individuals are more willing to discuss climate change when they are made aware of others' concern. Similar interventions have been shown to be effective at increasing charitable donations (Frey and Meier, 2004) and public good contributions (Ostrom, 2000), as well as moving voters to more moderate political positions (Ahler, 2014).

Whenever the pro-climate behaviour corresponds to the state of the world, it can be more effective to disseminate publicly which type is more likely held by the majority (Section 4.5). Whenever a policy maker knows or can influence what type of player moves first, a rationality boost (Section 4.4) can have a similar effect to information dissemination about the distribution of types.

Chapter 5

Conclusion

Information asymmetries are a salient factor in determining individuals' incentives – with profound implications for the outcomes in economic exchange. Therefore, developing a deeper understanding of the value of information for decision-makers has long played a central role in economic theory, including the development of a general framework of Bayesian games for applied economic modelling of information asymmetries, a formalisation of concepts like adverse selection or moral hazard, as well as tools such as screening and signalling.

In this thesis, I built on these longstanding concepts in microeconomic theory and information economics. I combined those with behavioural elements to examine real-world phenomena through a different lens. Specifically, by introducing *non-standard preferences*, *non-standard beliefs*, and *non-standard decision-making* into the standard models of signalling (Spence, 1973), Lancaster's (1966)'s product characteristics approach, and the theoretical framework of preference falsification (Kuran, 1987a,b, 1995), I contributed to a way of doing economics that is relevant to society.

In the **first** essay, I developed a game-theoretic framework in which an evaluator is biased against females and his bias is common knowledge. I showed that, above a threshold bias, the evaluator prefers a blind audition to provide high effort incentives exclusively for high-ability applicants. I also showed that a highly biased evaluator's preferences align with those of a highly able female. The introduction of performance uncertainty may lead to market failure or may render informed auditions more profitable, rationalising ability-targeting interventions. While a full employment subsidy can be a viable alternative to a blind audition for low biases, a simple quota cannot provide equal opportunity and targeted effort incentives in my benchmark model. Finally, I showed that it can be optimal to reveal ability through a gender-blind CV.

The framework I developed in this essay opens up intriguing avenues to further examine the potential trade-off between policies mitigating biases and effort incentives for applicants. For example, I assumed throughout that the evaluator's level of bias is common knowledge.

However, applicants may be uncertain or may underestimate the evaluator's bias. Moreover, the evaluator himself may only be partially sophisticated or naive about his bias when choosing the audition form. Besides the limitation that the framework currently cannot capture forms of bias that work at a subconscious level, I used a reduced-form bias parameter throughout. Hence, a promising avenue for research is to microfound the evaluator's bias. Another important avenue for further research is to introduce competition among applicants, for example, through a tournament model. The extension to competition among applicants would also provide an opportunity to compare blind auditions to affirmative action policies more broadly. For example, one could consider "plus factoring" as a mandated tie-breaking rule in a point-system for an informed audition that requires that, if two applicants came along and both had the same qualifications and experience, the evaluator would choose the female applicant as part of the affirmative action policy (Cornell Law School LII, n.d.).

In the **second** essay, I developed a model that explains a handmade strategy of firms in an era of technological advancement through a handmade effect on the consumer side. I deployed the model to generate insights whenever a firm chooses between a conventional production process and an alternative production process – including Fair Trade, organic produce, and local produce. I showed that when consumers are willing to pay a sufficiently high handmade premium, the monopolist chooses production by hand over superior machine production. When the firm is part of a duopoly, the existence of "conscious" shoppers can explain the firms' specialisation and, thus, the observed co-existence of handmade and machine-made products in the economy. Finally, through the introduction of shoppers uncertain about the appropriate consumption behaviour, a monopolist may be able to use a handmade strategy to signal a social norm of conscious consumption.

As I made several simplifying assumptions to model the "handmade" phenomenon, the essay opens up intriguing avenues to examine in greater detail a situation in which conscious shoppers and quality shoppers interact. For example, I assumed that quality shoppers gain utility only from product quality; they are not concerned about their reputation when making the buying decision. More realistically, however, while not caring about the production process per se, quality shoppers may incur a positive spillover when buying a handmade product that is also purchased by conscious shoppers. One could also envisage a situation in which regulation forces firms to state their production process on the packaging. This would modify my Assumption 2 to a "two-sided concern" as both hand and machine production would be salient to the conscious shopper. Moreover, in the baseline model, I limit my firm to producing a single type of product – handmade or machine-made. Given the significance of this assumed restriction, I already devoted Section 3.3.2 to a robustness check that analysed the monopolist's production decision when not limited to producing only one type of product. However, one could study more fully how relaxing this restriction would affect, say, the duopoly or the signalling game. Finally, my additional materials in Section A.2 provide a starting point to: (i) further study other equilibrium candidates in the class of subgames in which the duopolists use different production processes (i.e., specialise), as well as sequential moves in the duopoly, which is arguably a better fit for many applications where firms are aware of whether their competitors are engaging in handmade production; and (ii) consider more extensively pooling and hybrid equilibria in the signalling game, as well as a specification where conformists' willingness to pay is increasing smoothly in the expected fraction of conscious shoppers.

In the **third** essay, we analysed heuristic decision-making in social settings without communication. To understand its effect on individual behaviour, we first derived the conditions under which a credulous, sceptical or rational player falsifies her type. To understand the effect on aggregate behaviour and answer our main research question, we then characterised the mapping between types and actions in a group of four players. Finally, we considered the implications of making the state of the world common knowledge.

Throughout the analysis, we assumed that players are aware of their type. However, similar to the first essay, due to dynamic inconsistency, context dependency and framing effects, this need not be the case. Thus, the inclusion of uncertainty about one's type is a promising direction for future research to fully understand the effects and potential interdependencies of these two types of cognitive limitations (i.e., attribution bias and limited recall). Finally, my additional materials in Section A.3.2 provide a starting point to further explore an alternative model of scepticism that places a sceptical player on a continuum between a fully rational and credulous player. This may provide a more intuitive definition of sophistication where a sophisticated player i believes that there are two kinds of players in the population (i.e., individualistic and moderately conformist players) when, in fact, all players are conformist to some degree. Another restriction in the baseline model is that the groups considered are homogeneous in their level of rationality. Given the significance of this assumed restriction, we already devoted Section A.3.3 to a generalisation of four-player groups featuring any combination of rationality levels. A starting point for examining in more depth the effect of heterogeneity in the level of rationality may be the alternative model of scepticism set out in Section A.3.2: by varying the probability with which a fellow player is believed to be only concerned with matching their type amongst sceptical players 2 and 3 (i.e., ε_2 and ε_3), we have a mixture of rationality levels within the four-player group. This highlights that in this alternative model of sophistication we would need to address higher-order beliefs by assuming ε_i to be common knowledge.

Appendix A

Additional Materials

A.1 Chapter 1 Additional Materials

A.1.1 Refinements in Separating Blind Audition

For the evaluator's belief that only high-ability applicants exert high effort to be consistent, he needs to have a sufficiently high bias; that is, $\beta \in \beta_B^H := (\sqrt{17} - 3, 2)$. In this case, low-ability applicants do not participate and the information set q_L is not reached. Therefore, the evaluator's belief at q_L is not determined by equilibrium play; in other words, perfect Bayesian equilibrium does not place restrictions on this belief as it is off the equilibrium path.

Furthermore, note that the Intuitive Criterion (Cho and Kreps, 1987) is moot. In the separating blind audition, exerting low effort is equilibrium dominated for all types. In particular, for a low-ability applicant, the equilibrium payoff from dropping out is greater than the highest possible payoff from exerting low effort. For a high-ability applicant, the equilibrium payoff from exerting high effort is greater than the highest possible payoff from exerting low effort. Therefore, the requirement that the evaluator's belief at q_L places zero probability on nodes that are reached only if an applicant plays an equilibrium dominated strategy does not apply. It is not possible for the belief at q_L to place zero probability on all four nodes simultaneously.

A.1.1.1 Refinement based on Deviation Payoffs

I argue, however, that the evaluator's belief that the applicant is of low ability when observing a deviation to q_L is a reasonable restriction. Suppose to the contrary that, for a given bias $\beta \in \beta_B^H$, in the separating blind audition, the evaluator also believes the applicant to be of high ability if she delivers a low-quality performance behind the curtain. This essentially means that, if a low-quality performance is to be observed, it is because a high-ability applicant has trembled and chosen an easy piece by accident, or she rests on her laurels of being believed to be of high ability irrespective of the difficulty of the piece performed. In other words, a low-quality performance is not observed because a low-ability applicant is attracted to the audition to exploit the evaluator's belief at q_L. This perturbation of the evaluator's belief implies that, he expects a higher gross utility of

$$\mathbb{E}[V|q_L] = q_L + \mathbb{E}[\eta|q_L] - \frac{\beta}{2} = 1 + 2 - \frac{\beta}{2}$$

from hiring the applicant at q_L. Therefore, the probability of being hired at this information set changes to

$$\Pr(\text{hired}|q_L) = \Pr\left(\bar{U} \leqslant 3 - \frac{\beta}{2}\right) = \frac{2+\beta}{4+2\beta}.$$

Note that the hiring probability at q_L is higher than the one under the evaluator's original belief, $\frac{\beta}{4+2\beta}$. Intuitively, if the evaluator believes the applicant behind the curtain to be of high ability when hearing an easy piece, he should be more willing to hire her and overlook her mistake. Due to the altered hiring probability, the applicant's effort incentives change to

$$\begin{aligned}
U(e_L, \eta_L) &= \frac{2+\beta}{4+2\beta}w - \frac{1}{2}, \\
U(e_H, \eta_L) &= \frac{4+\beta}{4+2\beta}w - 2, \\
U(e_L, \eta_H) &= \frac{2+\beta}{4+2\beta}w - \frac{1}{4}, \\
U(e_H, \eta_H) &= \frac{4+\beta}{4+2\beta}w - 1,
\end{aligned}$$

where $w := \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$.

First, consider a low-ability applicant under the perturbed belief at q_L. Before, this type was deterred from participating for $\beta \in \beta_B^H$. Now, exerting low effort is worthwhile for all biases in this interval (Figure A.1a). The low-ability applicant is, in fact, exploiting the evaluator's belief; it plays out in her favour. For $\beta \in \beta_B^H$, the utility gain from deviating from O to e_L is given by

$$\big[U(e_L, \eta_L) - U(O, \eta_L)\big] = 1 - \frac{\beta}{4},$$

which is monotonically decreasing in the bias for all possible values. Therefore, the lowest possible utility gain from deviating to e_L occurs as β approaches its maximum:

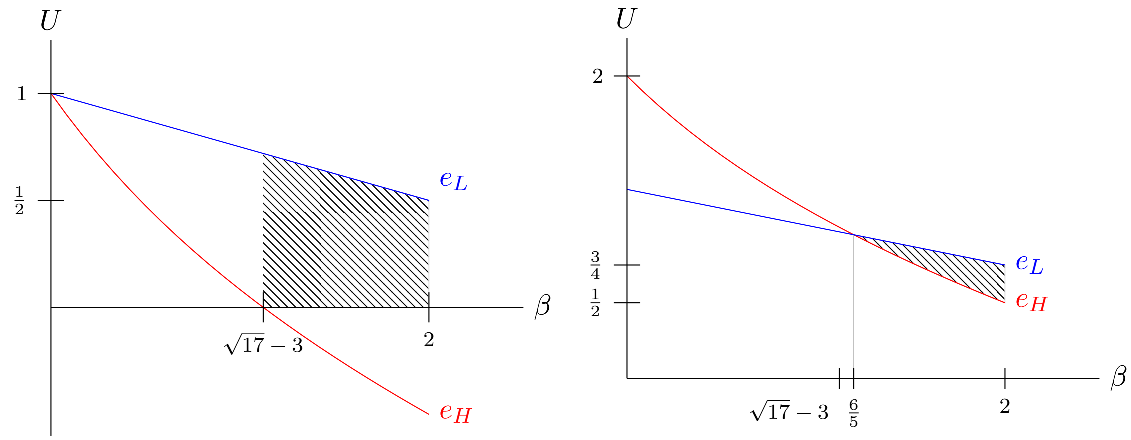
$$\lim_{\beta \uparrow 2} \big[U(e_L, \eta_L) - U(O, \eta_L)\big] = \frac{1}{2}. \tag{A.1}$$

Second, consider a high-ability applicant under the perturbed belief at q_L. Before, this type was exerting high effort for $\beta \in \beta_B^H$. Now, exerting high effort is only worthwhile if the bias does not exceed $\frac{6}{5}$. Exerting low effort is optimal for any bias exceeding this threshold (Figure A.1b). The high-ability applicant essentially avoids the higher cost of effort and rests on her laurels by preparing an easy piece. For $\beta \in \beta_B^H$, the utility gain from deviating from e_H to e_L is given by

$$\left[U(e_L, \eta_H) - U(e_H, \eta_H)\right] = \frac{5}{4} - \frac{4}{2+\beta},$$

which is monotonically increasing in the bias for all possible values. Therefore, the greatest possible utility gain from deviating to e_L occurs as β approaches its maximum:

$$\lim_{\beta \uparrow 2} \left[U(e_L, \eta_H) - U(e_H, \eta_H)\right] = \frac{1}{4}. \tag{A.2}$$



(a) Low-ability applicant. (b) High-ability applicant.

Figure A.1: Effort choices in a separating blind audition.

Finally, observe that for $\beta \in (\frac{6}{5}, 2)$, both high- and low-ability applicants gain from deviating to low effort under the perturbed belief. The utility gain for low-ability from deviating, however, is always greater than the utility gain for high-ability applicants from deviating. This follows immediately because the lowest possible utility gain for the former in equation (A.1) exceeds the largest possible utility gain for the latter in equation (A.2). As a result, the evaluator should say to himself that the applicant behind the curtain is more likely to be a low-ability applicant exploiting his belief. Put differently, a high-ability applicant trembling to an easy piece or resting on her laurels of being believed to be of high ability irrespective of the difficulty of the piece is less likely.

Observe also that for $\beta \in (\sqrt{17}-3, \frac{6}{5}]$, only low-ability applicants have an incentive to deviate to low effort under the perturbed belief. High-ability applicants exert high effort. For this range of bias, by the Intuitive Criterion, the evaluator should believe the applicant to be of low-ability with probability one when observing a low-quality performance.

A.1.1.2 Refinement based on D1-Criterion

Under D1, similar to the Intuitive Criterion, deviations of the applicant emerge as the consequence of a rational decision. D1, however, is capable to put more structure on the evaluator's belief at q_L in the separating blind audition: it restricts the off-path belief to be a point belief with all mass on the type who is most likely to deviate.

Define $D(\theta, T, e_L)$ to be the set of the evaluator's mixed strategy best responses to action e_L and beliefs concentrated on T that make type θ strictly prefer e_L to her equilibrium strategy. Let $D^0(\theta, T, e_L)$ be the set of mixed best responses that make type θ exactly indifferent.

Definition 1. A type θ is deleted for strategy e_L under the D1-Criterion if there is a θ' such that

$$\{D(\theta, \Theta, e_L) \cup D^0(\theta, \Theta, e_L)\} \subset D(\theta', \Theta, e_L).$$

From Definition 1 follows that, if the set of the evaluator's mixed best responses (that is, the range of hiring probabilities) that make high-ability applicants willing to deviate to low effort is strictly smaller than the set of best responses that make low-ability applicants willing to deviate, then the evaluator should believe that low-ability applicants are infinitely more likely to deviate to low effort than their high-ability counterparts are. In other words, D1 tests a deviation to low effort for a given type with respect to each particular mixed best response of the evaluator (Fudenberg and Tirole, 1991, p.451-452).

Proposition 1. *After observing q_L, the set of the evaluator's best responses improving the expected equilibrium utility of low-ability applicants is a superset of the set of best responses improving the expected equilibrium utility of high-ability applicants. Therefore, the evaluator should infer that he deals with the former and put zero weight on the latter.*

Proof of Proposition 1. First, consider low-ability males and females in the separating blind audition. The minimal hiring probability at the low-performance information set that induces these types to deviate from O to e_L is given by

$$\Pr(\text{hired}|q_L)w - \frac{1}{2} \geqslant 0,$$
$$\Rightarrow \Pr(\text{hired}|q_L) \geqslant \frac{1}{6-\beta}. \tag{A.3}$$

Second, consider high-ability males and females in the separating blind audition. The minimal hiring probability at the low-performance information set that induces these types to deviate from e_H to e_L is given by

$$\Pr(\text{hired}|q_L)w - \frac{1}{4} \geqslant \Pr(\text{hired}|q_H)w - 1,$$

$$\Rightarrow \Pr(\text{hired}|q_L) \geqslant \frac{4+\beta}{4+2\beta} - \frac{3}{12-2\beta}, \tag{A.4}$$

where the probability of being hired at the high-performance information set is fixed at its equilibrium value, $\frac{4+\beta}{4+2\beta}$.

Finally, for a given bias, compare inequality (A.3) with inequality (A.4) to conclude that

$$\{D((\eta_H,\cdot),\Theta,e_L) \cup D^0((\eta_H,\cdot),\Theta,e_L)\} \subset D((\eta_L,\cdot),\Theta,e_L).$$

□

Figure A.2 shows graphically that the set of the evaluator's mixed best responses to e_L that make low-ability applicants prefer low effort to their equilibrium strategy (light and dark grey area) strictly contains the set of mixed best responses for their high-ability counterparts (dark grey area) for all possible values of bias. By Definition 1, therefore, a deviation to low effort must be interpreted as coming from low-ability applicants.

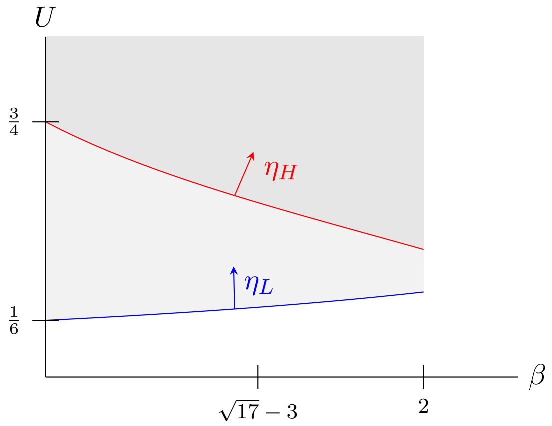


Figure A.2: Set of the evaluator's mixed best responses to low effort inducing a deviation of low- and high-ability applicants.

A.1.2 Solution under Performance Uncertainty

A.1.2.1 Informed Audition under Performance Uncertainty

The introduction of performance uncertainty changes the ex-ante hiring probabilities when effort is low: they are a convex combination of low effort mapping into low performance quality as well as the ε chance of attaining high performance quality as a consequence of randomness. Intuitively, such performance uncertainty makes exerting low effort relatively more attractive and alters the applicant's participation and incentive constraints. That, in turn, affects the applicant's optimal effort choice in the second stage as well as the partitioning of the evaluator's bias in Lemma 1. I define $\bar{\beta}_{LM}(\varepsilon) := \{\beta : U(e_L, \eta_L, m) = 0, \beta \geqslant 0\}$ and $\bar{\beta}_{HA}(\varepsilon) := \{\beta : U(e_L, \eta_H, \cdot) = U(e_H, \eta_H, \cdot)\}$.

Lemma 1′. *Under Assumption 5, for $\varepsilon < \bar{\varepsilon} := \frac{1}{3}$, the applicants' effort responses partition the evaluator's bias $\beta \in [0, 2]$ into three regions: (i) For a low bias, $\beta \in \beta_I^L(\varepsilon) := [0, \bar{\beta}_{LM}(\varepsilon))$, only high-ability applicants participate and exert high effort. (ii) For a moderate bias, $\beta \in \beta_I^M(\varepsilon) := [\bar{\beta}_{LM}(\varepsilon), \bar{\beta}_{HA}(\varepsilon)]$, low-ability males also participate and exert low effort. (iii) For a high bias, $\beta \in \beta_I^H(\varepsilon) := (\bar{\beta}_{HA}(\varepsilon), 2]$, all applicants except low-ability females participate and exert low effort.*

Proof of Lemma 1′. First, note that, when taking into account the performance uncertainty in the mapping from low effort to performance quality, the expected utilities of the four types of applicants are

$$U(e_L, \eta_L, f) = \frac{\varepsilon}{2+\beta} w - \frac{1}{2}, \qquad U(e_L, \eta_L, m) = \left[\frac{\beta(1-\varepsilon)}{2+\beta} + \frac{(1+\beta)\varepsilon}{2+\beta}\right] w - \frac{1}{2},$$
$$U(e_H, \eta_L, f) = \frac{1}{2+\beta} w - 2, \qquad U(e_H, \eta_L, m) = \frac{1+\beta}{2+\beta} w - 2,$$
$$U(e_L, \eta_H, f) = \left[\frac{1-\varepsilon}{2+\beta} + \frac{2\varepsilon}{2+\beta}\right] w - \frac{1}{4}, \quad U(e_L, \eta_H, m) = \left[\frac{(1+\beta)(1-\varepsilon)}{2+\beta} + \varepsilon\right] w - \frac{1}{4},$$
$$U(e_H, \eta_H, f) = \frac{2}{2+\beta} w - 1, \qquad U(e_H, \eta_H, m) = w - 1$$

in an informed audition, where $w := \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. Next, I consider the effect of performance uncertainty on incentive compatibility and participation constraints for all four types of applicants.

For low-ability females, a combination of high performance uncertainty and low bias makes participating and exerting low effort in an informed audition worthwhile. In particular, low-ability females exert low effort instead of dropping out if

$$U(e_L, \eta_L, f) \geqslant 0 \Rightarrow 0 \leqslant \beta \leqslant \frac{2(3\varepsilon - 1)}{\varepsilon + 1}.$$

Given that the bias is constrained to be non-negative, this results in the threshold $\bar{\beta}_{LF}(\varepsilon) := \frac{2(3\varepsilon-1)}{\varepsilon+1}$ for $\varepsilon \in [\frac{1}{3}, \frac{1}{2})$ and $\bar{\beta}_{LF} := 0$ for $\varepsilon \in [0, \frac{1}{3})$. At $\bar{\varepsilon} := \frac{1}{3}$, low effort is worthwhile if the evaluator is impartial. Beyond this critical value, the threshold bias for participation is increasing in ε:

$$\frac{d\bar{\beta}_{LF}(\varepsilon)}{d\varepsilon} = \frac{8}{(\varepsilon+1)^2} > 0 \ \forall \ \varepsilon \in \left[\frac{1}{3}, \frac{1}{2}\right).$$

In other words, beyond a critical level of performance uncertainty, the evaluator can be increasingly biased against female applicants and still make low effort worthwhile. This is depicted in Figure A.3a, where $\varepsilon_0 = 0$.

In contrast, performance uncertainty has an adverse effect on high-ability females. In particular, high-ability females exert low rather than high effort if

$$U(e_L, \eta_H, f) \geqslant U(e_H, \eta_H, f) \Rightarrow \beta \geqslant \frac{6 - 12\varepsilon}{5 - 2\varepsilon}.$$

Thus, $\bar{\beta}_{HF}(\varepsilon) := \frac{6-12\varepsilon}{5-2\varepsilon}$ for $\varepsilon \in [0, \frac{1}{2})$. For all levels of performance uncertainty, this threshold bias is decreasing in ε:

$$\frac{d\bar{\beta}_{HF}(\varepsilon)}{d\varepsilon} = -\frac{48}{(5-2\varepsilon)^2} < 0 \;\forall\; \varepsilon \in \left[0, \frac{1}{2}\right).$$

In fact, as $\varepsilon \uparrow \frac{1}{2}$, the threshold bias approaches zero (Figure A.3b). Intuitively, the more uncertain the environment, the lower the bias has to be to distort effort incentives for a highly able female.

For low-ability males, the threshold bias, beyond which participating and exerting low effort is worthwhile, falls with the introduction of performance uncertainty to the environment. In particular, low-ability males exert low effort instead of dropping out if

$$U(e_L, \eta_L, m) \geqslant 0 \Rightarrow \beta \geqslant \frac{5-\varepsilon-\sqrt{\varepsilon^2+14\varepsilon+17}}{2} \geqslant 0.$$

Thus, $\bar{\beta}_{LM}(\varepsilon) := \frac{5-\varepsilon-\sqrt{\varepsilon^2+14\varepsilon+17}}{2}$ for $\varepsilon \in [0, \frac{1}{3})$ and $\bar{\beta}_{LM} := 0$ for $\varepsilon \in [\frac{1}{3}, \frac{1}{2})$. For $\varepsilon < \bar{\varepsilon}$, this threshold bias is decreasing in ε:

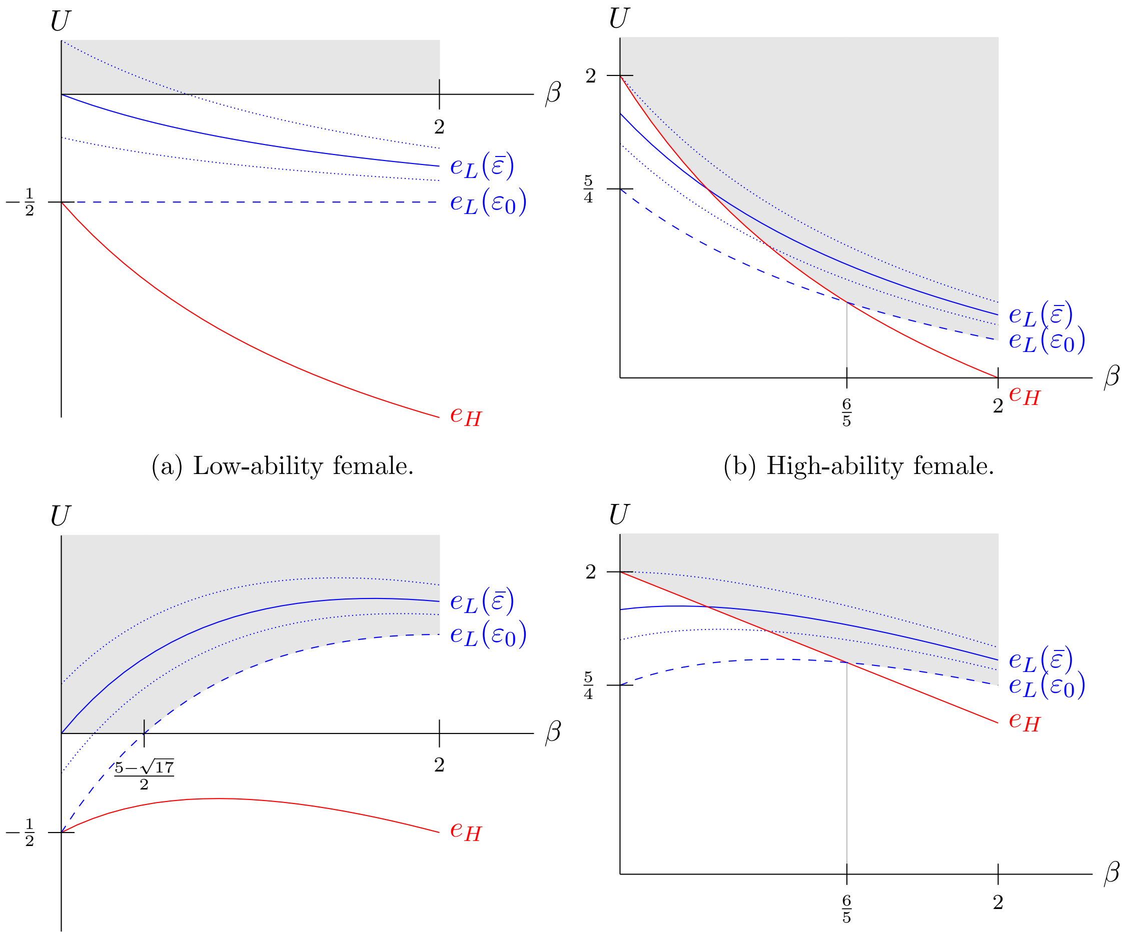
$$\frac{d\bar{\beta}_{LM}(\varepsilon)}{d\varepsilon} = -\frac{1}{2}\left[1 + \frac{\varepsilon+7}{\sqrt{\varepsilon^2+14\varepsilon+17}}\right] < 0 \;\forall\; \varepsilon \in \left[0, \frac{1}{3}\right).$$

For $\varepsilon \geqslant \bar{\varepsilon}$, low-ability males are always induced to participate and exert low effort even if the evaluator is impartial (Figure A.3c).

Performance uncertainty has an adverse effect on high-ability males. In particular, high-ability males exert low rather than high effort if

$$U(e_L, \eta_H, m) \geqslant U(e_H, \eta_H, m) \Rightarrow \beta \geqslant \frac{6-12\varepsilon}{5-2\varepsilon}.$$

Thus, $\bar{\beta}_{HM}(\varepsilon) := \frac{6-12\varepsilon}{5-2\varepsilon}$ for $\varepsilon \in [0, \frac{1}{2})$. Because this threshold bias coincides with the one derived for high-ability females, it follows immediately that $\bar{\beta}_{HM}(\varepsilon)$ is decreasing in ε. Similarly, as $\varepsilon \uparrow \frac{1}{2}$, the threshold bias approaches zero (Figure A.3d). I define $\bar{\beta}_{HA}(\varepsilon) := \bar{\beta}_{HF}(\varepsilon) = \bar{\beta}_{HM}(\varepsilon)$ to highlight that the thresholds for high-ability females and males coincide.



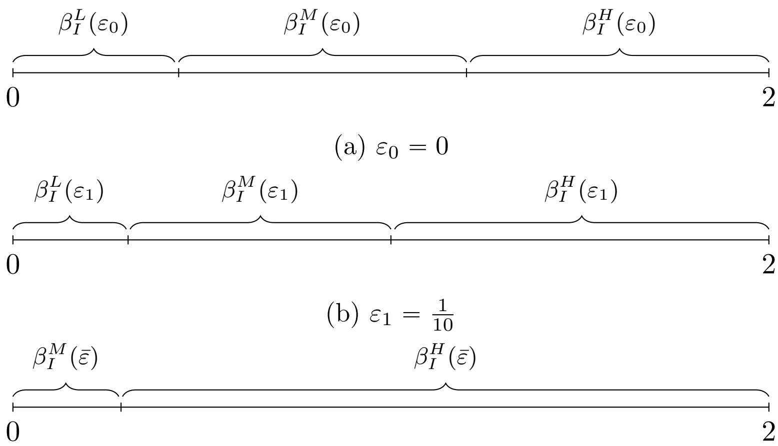
(a) Low-ability female. (b) High-ability female.

(c) Low-ability male. (d) High-ability male.

Figure A.3: Effort choices in an informed audition under performance uncertainty.

□

To get some intuition behind Lemma 1′; that is, how the applicants' effort responses under rising performance uncertainty partition the evaluator's bias, fix $\varepsilon_0 = 0$ and $\varepsilon_1 = \frac{1}{10}$. First, Figure A.4 illustrates that the interval $\beta_I^H(\varepsilon)$ is expanding to the left as the degree of performance uncertainty increases. Second, an increase in performance uncertainty causes the interval $\beta_I^L(\varepsilon)$ to shrink at a faster rate than the interval $\beta_I^M(\varepsilon)$. As a result, $\bar{\varepsilon}$ marks the threshold beyond which $\beta_I^L(\varepsilon)$ vanishes. These two observations are intuitive: the more uncertain the environment, the lower the bias of the evaluator needs to be to distort effort incentives for the highly able. This is exacerbated by the fact that performance uncertainty attracts low-ability applicants. The threshold $\bar{\varepsilon}$ determines which gender benefits from rising performance uncertainty: as the degree of performance uncertainty approaches the threshold from below, low-ability males benefit and participate for a greater range of biases, with no change for low-ability females. Beyond this threshold, as the degree of performance uncertainty approaches its supremum, low-ability females benefit and participate more often.



(a) $\varepsilon_0 = 0$

(b) $\varepsilon_1 = \frac{1}{10}$

(c) $\bar{\varepsilon} = \frac{1}{3}$

Figure A.4: Partition of the evaluator's bias for different degrees of performance uncertainty.

In what follows, I assume $\varepsilon < \bar{\varepsilon}$. While this restriction is not without loss of generality, it ensures that low-ability females never participate and that the results are comparable to those derived under Assumption 4.

Proposition 1′. *Suppose Assumption 5 holds and $\varepsilon < \bar{\varepsilon}$. (i) Suppose further that $\beta \in \beta_I^L(\varepsilon)$ where $\beta_I^L(\varepsilon)$ is defined as in Lemma 1′(i). Then, the evaluator's expected net utility (2.4) is unchanged. (ii) Suppose $\beta \in \beta_I^M(\varepsilon)$ where $\beta_I^M(\varepsilon)$ is defined as in Lemma 1′(ii). Then, $\mathbb{E}[\Pi_I|\beta \in \beta_I^M(\varepsilon)] \gtrless 0$. (iii) Suppose $\beta \in \beta_I^H(\varepsilon)$ where $\beta_I^H(\varepsilon)$ is defined as in Lemma 1′(iii). Then, $\mathbb{E}[\Pi_I|\beta \in \beta_I^H(\varepsilon)] \gtrless 0$.*

Proof of Proposition 1′. Suppose the evaluator's bias against female applicants is low; that is, $\beta \in \beta_I^L(\varepsilon)$ where $\beta_I^L(\varepsilon)$ is defined as in Lemma 1′(i). Then, the evaluator's expected net utility (2.4) is not affected by performance uncertainty because the bias-uncertainty combination is sufficiently low as to provide targeted effort incentives.

The proof for a moderate and high bias follows the steps detailed in the proof of Proposition 1 when there is no performance uncertainty. In particular, suppose the evaluator is moderately biased against female applicants; that is, $\beta \in \beta_I^M(\varepsilon)$ where β_I^M is defined as in Lemma 1′(ii). Then,

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^M(\varepsilon)] = \frac{1-p}{2}\left[(1-\varepsilon)\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right) + \varepsilon\frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)\right] + \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right) + \left(1+\frac{\beta}{2}\right)\right] \gtrless 0.$$

Suppose the evaluator is highly biased against female applicants; that is, $\beta \in \beta_I^H(\varepsilon)$ where β_I^H is defined as in Lemma 1′(iii). Then,

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^H(\varepsilon)] = \frac{1-p}{2}\left[(1-\varepsilon)\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)+\varepsilon\frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)\right]$$
$$+\frac{p}{2}\left[(1-\varepsilon)\frac{1}{2+\beta}\left(-\frac{\beta}{2}\right)+\varepsilon\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right)+(1-\varepsilon)\frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)+\varepsilon\left(1+\frac{\beta}{2}\right)\right] \gtrless 0.$$

□

Proposition 1′ highlights that performance uncertainty alters the evaluator's expected net utility if an informed audition cannot provide targeted effort incentives; that is, if requiring a CV does not pose an effective means to screen applicants by ability.

A.1.2.2 Blind Audition under Performance Uncertainty

I now consider the effect of performance uncertainty on the pooling and separating equilibrium analysed in Section 2.4. A key question for policy is whether the separating equilibrium continues to exist and remains more profitable for the evaluator. This would ensure that my predictions generalise to an uncertain environment and that a strategy of commitment to no information can be beneficial if the evaluator knows that he would otherwise not be impartial.

Pooling Equilibrium Suppose the evaluator's bias and the degree of performance uncertainty are sufficiently low to induce all applicants to exert high effort. Given the asymmetry in Assumption 5, low performance quality continues to be off the equilibrium path. As in Section 2.4, I assume that the evaluator believes the applicant to be of low ability when observing a low-quality performance.

I now show that under performance uncertainty p and ε usually determine jointly whether a high-effort pooling equilibrium can be supported and that they place an upper bound on the evaluator's bias in any such pooling equilibrium. I define $\bar{\beta}_{pool}(p,\varepsilon) :=$ $\min\{\beta(p), \beta(p,\varepsilon)\}$ where $\beta(p,\varepsilon) := \{\beta : U(e_H,\eta_L) = U(e_L,\eta_L)\}$.

Lemma 2′. *Under Assumption 5, (i) for $p < \frac{1}{3}$, no high-effort pooling equilibrium exists. (ii) For $p = \frac{1}{3}$, a high-effort pooling equilibrium exists if the evaluator is unbiased and $\varepsilon \leqslant \frac{1}{4}$. (iii) For $\frac{1}{3} < p < 1$, a high-effort pooling equilibrium exists if $\beta \in \beta_B^L(\varepsilon) := [0, \bar{\beta}_{pool}(p,\varepsilon)]$ and $\bar{\beta}_{pool}(p,\varepsilon) > 0$.*

Proof of Lemma 2′. First, note that, when taking into account the performance uncertainty in the mapping from low effort to performance quality, the expected utilities are

$$U(e_L,\eta_L) = \left[\frac{\beta}{4+2\beta}(1-\varepsilon)+\frac{(2(1+p)+\beta)}{4+2\beta}\varepsilon\right]w-\frac{1}{2},$$

$$
\begin{aligned}
U(e_H,\eta_L) &= \frac{2(1+p)+\beta}{4+2\beta}w-2,\\
U(e_L,\eta_H) &= \left[\frac{\beta}{4+2\beta}(1-\varepsilon)+\frac{(2(1+p)+\beta)}{4+2\beta}\varepsilon\right]w-\frac{1}{4},\\
U(e_H,\eta_H) &= \frac{2(1+p)+\beta}{4+2\beta}w-1
\end{aligned}
$$

in a pooling blind audition, where $w := \mathbb{E}[\bar{U}] = 3-\frac{\beta}{2}$. I now consider the three cases separately.

(i) First, note that $U(e_H,\eta_L)$ is independent of ε. Therefore, by Lemma 2(i), for $p<\frac{1}{3}$, pooling on high effort is never possible.

(ii) For $p=\frac{1}{3}$, $U(e_H,\eta_L)=0$ at $\beta=0$. By Lemma 2(i), $U(e_H,\eta_L)<0$ for all $\beta\in(0,2]$. This implies that the low-ability applicant is indifferent between exerting high effort and not participating only if the evaluator is unbiased. Under performance uncertainty, I also need to ensure incentive compatibility. Because it is optimal for high-ability applicants to exert high effort whenever it is optimal for low-ability applicants to exert high effort, this constraint is given by

$$
U(e_H,\eta_L)\Big|_{\beta=0,p=\frac{1}{3}} \geqslant U(e_L,\eta_L)\Big|_{\beta=0,p=\frac{1}{3}} \Rightarrow \varepsilon\leqslant\frac{1}{4}.
$$

Thus, pooling on high effort is possible at $p=\frac{1}{3}$ if $\beta=0$ and $\varepsilon\leqslant\frac{1}{4}$.

(iii) For $\frac{1}{3}<p<1$, for low degrees of performance uncertainty, the upper bound on β for which a pooling equilibrium can be supported is determined by the indifference condition

$$
U(e_H,\eta_L)=\frac{2(1+p)+\beta}{4+2\beta}\left(3-\frac{\beta}{2}\right)-2=0.
$$

Given that $\beta\in[0,2]$, the above equation is uniquely solved by $\beta(p):=\sqrt{p^2+16p}-p-2$. For larger degrees of performance uncertainty, the upper bound on β for which a pooling equilibrium can be supported is determined by the indifference condition

$$
U(e_H,\eta_L)=U(e_L,\eta_L).
$$

Given that $\beta\in[0,2]$, $\varepsilon\in[0,\frac{1}{2})$ and $p\in(0,1)$, the above equation is uniquely solved by $\beta(p,\varepsilon):=\frac{6(p\varepsilon-p+\varepsilon)}{p\varepsilon-p+\varepsilon-4}$. Therefore, the upper bound on a high-effort pooling equilibrium is given by $\bar{\beta}_{pool}(p,\varepsilon):=\min\{\beta(p),\beta(p,\varepsilon)\}$, provided that $\bar{\beta}_{pool}(p,\varepsilon)$ is non-negative. Otherwise, no high-effort pooling equilibrium exists. □

Pooling on high effort is possible if the prior that the applicant is of high ability is sufficiently large and the evaluator's bias as well as the degree of performance uncertainty is not too extreme. In particular, for a fixed prior $p>\frac{1}{3}$, once a critical level of performance uncertainty is reached, the range of biases for which pooling on high effort is an equilibrium is decreasing in performance uncertainty. This is driven by the fact that low effort becomes relatively more attractive. Conversely, for a fixed degree of performance uncertainty $\varepsilon > 0$, once a critical level of bias is reached, the range of biases for which pooling on high effort is an equilibrium outcome is increasing in the prior that the applicant is of high ability. Intuitively, while the ability composition of the applicant pool affects the applicant's utility under either effort choice, the effect of an increasingly able applicant pool is less pronounced when she chooses to exert low effort. This is because the effect of a change in the ability composition of the applicant pool is scaled down by the performance uncertainty that arises from exerting low effort. Therefore, this comparative static is robust to the introduction of performance uncertainty: an increasingly able applicant pool makes it easier for for the fewer low-ability to hide behind the more-and-more high-ability applicants. If a high-effort pooling equilibrium exists under performance uncertainty, the evaluator's expected net utility (2.8) is unchanged.

Fully Separating Equilibrium Suppose the evaluator's bias is sufficiently high and the degree of performance uncertainty is suffficiently low to induce high-ability applicants to exert high effort and low-ability applicants to drop out.[1] Under performance uncertainty, this qualification, which I refer to as full separation, is important: it ensures that the evaluator has degenerate posterior beliefs at the information set q_H. In other words, upon observing a high-quality performance, the evaluator is certain to face a highly able applicant who is male or female with equal probability. If, in contrast, low-ability applicants were to exert low effort, the evaluator could not be certain about the applicant's ability at q_H due to the ε probability that low effort results in a high-quality performance. In this case, I would need to use Bayes' rule in deriving the evaluator's expected gross utility conditional upon observing a high-quality performance and this would alter the hiring probability at q_H (see Section A.1.2.2)

Lemma 3′. *Under Assumption 5, for $\varepsilon > \hat{\varepsilon}$, a fully separating equilibrium does not exist.*

Proof of Lemma 3′. First, note that when taking into account the performance uncertainty in the mapping from low effort to performance quality, the expected utilities are

$$\begin{aligned}
U(e_L, \eta_L) &= \left[\frac{\beta}{4+2\beta}(1-\epsilon) + \frac{4+\beta}{4+2\beta}\epsilon\right] w - \frac{1}{2} \\
U(e_H, \eta_L) &= \frac{4+\beta}{4+2\beta} w - 2 \\
U(e_L, \eta_H) &= \left[\frac{\beta}{4+2\beta}(1-\epsilon) + \frac{4+\beta}{4+2\beta}\epsilon\right] w - \frac{1}{4}
\end{aligned}$$

[1] As before, the evaluator believes the applicant to be of low ability at q_L.

$$U(e_H, \eta_H) = \frac{4+\beta}{4+2\beta} w - 1$$

in a fully separating blind audition, where $w := \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. In now show that, for $\varepsilon > \hat{\varepsilon}$, there is no range of biases that can support a separating equilibrium in which low-ability applicants do not participate.

First, note that low-ability applicants must find it optimal to drop out of the audition. This gives the lower bound $\underline{\beta}_{full}$ and the upper bound $\bar{\beta}_{full}(\varepsilon)$ for which a fully separating equilibrium can be supported. Given that $\frac{\partial U(e_H,\eta_L)}{\partial \beta} < 0$ and $\frac{\partial U(e_L,\eta_L)}{\partial \beta} > 0$, the lower bound is determined by the indifference condition

$$U(e_H, \eta_L) = 0.$$

For $\beta \in [0, 2]$, the above equation is uniquely solved by $\underline{\beta}_{full} = \sqrt{17} - 3$. The upper bound is determined by the indifference condition

$$U(e_L, \eta_L) = 0.$$

For $\beta \in [0, 2]$ and $\epsilon \in [0, \frac{1}{2})$, the above equation is uniquely solved by $\bar{\beta}_{full}(\varepsilon) = 2(1 - \varepsilon) - 2\sqrt{\varepsilon(4 + \varepsilon)}$. Therefore, for $\beta \in [0, \underline{\beta}_{full}]$, low-ability applicants exert high effort. For $\beta \in (\underline{\beta}_{full}, \bar{\beta}_{full}(\varepsilon))$, low-ability applicants do not participate and, for $\beta \in [\bar{\beta}_{full}(\varepsilon), 2]$, low-ability applicants exert low effort.

Second, note that the lower bound $\underline{\beta}_{full}$ is independent of ε whereas the upper bound is strictly decreasing in ε:

$$\frac{d\bar{\beta}_{full}(\varepsilon)}{d\varepsilon} = -2 - \frac{4 + 2\varepsilon}{\sqrt{\varepsilon(4+\varepsilon)}} < 0 \;\forall\; \varepsilon \in \left[0, \frac{1}{2}\right).$$

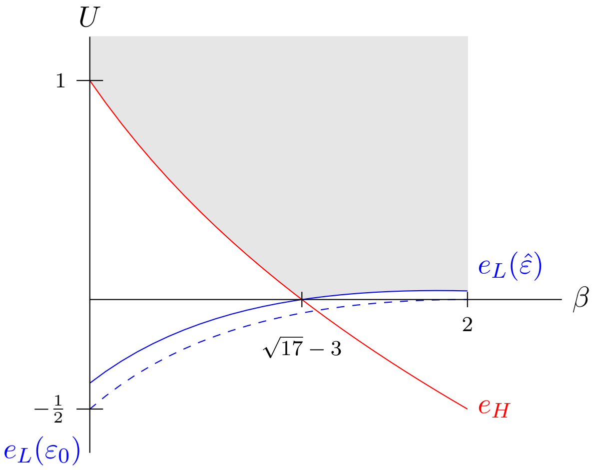
Therefore, the interval $(\underline{\beta}_{full}, \bar{\beta}_{full}(\varepsilon))$ in which low-ability applicants drop out of the audition is strictly decreasing in ε. In particular, if performance uncertainty exceeds a threshold, the interval is a zero-measure set: the upper bound falls below the lower bound and there is no range of biases that can support a fully separating equilibrium. This performance uncertainty treshold solves

$$\underline{\beta}_{full} := \sqrt{17} - 3 = 2(1 - \hat{\varepsilon}) - 2\sqrt{\hat{\varepsilon}(4 + \hat{\varepsilon})} =: \bar{\beta}_{full}(\hat{\varepsilon}),$$

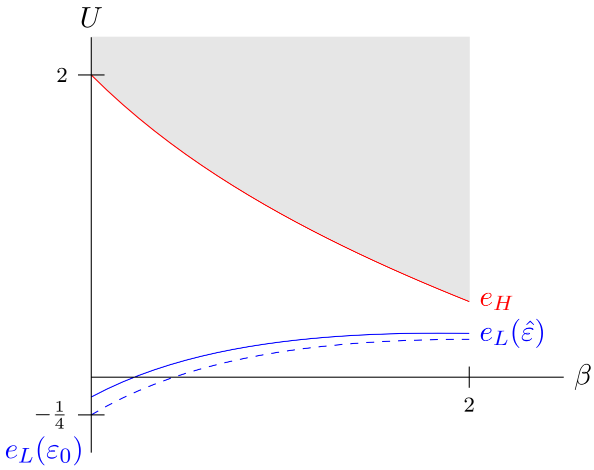
and is uniquely given by $\hat{\varepsilon} = \frac{13 - 3\sqrt{17}}{16}$.

Finally, note that I can neglect high-ability applicants in the analysis. Their participation is guaranteed: $U(e_H, \eta_H) > 0$. Moreover, for all $\varepsilon \in [0, \frac{1}{2})$, the upper bound $\beta(\varepsilon) = \frac{6(4\varepsilon - 3)}{4\varepsilon - 7}$ implied by the incentive compatibility constraint of high-ability applicants, $U(e_H, \eta_H) = U(e_L, \eta_H)$, is less stringent than the upper bound $\bar{\beta}_{full}(\varepsilon)$ implied by the participation constraint of low-ability applicants. □

My non-existence result is illustrated in Figure A.5a: performance uncertainty which exceeds a certain degree makes participating and exerting low effort for low-ability applicants worthwhile. More precisely, beyond the threshold $\hat{\varepsilon} := \frac{13-3\sqrt{17}}{16} \approx \frac{1}{25}$ (solid blue line), there is no range of biases that can support a separating equilibrium in which low-ability applicants do not participate. The frailty of the fully separating equilibrium to the introduction of performance uncertainty above this threshold stems from the fact that the payoff from exerting low effort for the less able is just below zero. As a result, the deterrence effect of the evaluator's off-path belief is voided by the chance of attaining a high performance quality as a consequence of randomness if this chance is at least $\hat{\varepsilon}$.



(a) Low-ability applicant.



(b) High-ability applicant.

Figure A.5: Effort choices in a fully separating blind audition under performance uncertainty.

Partially Separating Equilibrium The frailty of the fully separating equilibrium to the introduction of performance uncertainty above a threshold suggests that there exists a less extreme separating equilibrium for a range of bias-prior-uncertainty combinations, in which highly able applicants exert high effort and the less able exert low effort. I refer to this outcome as partial separation. Note that partial separation was not an equilibrium under Assumption 4: I specified the evaluator's beliefs off the equilibrium path to be such that he believes the applicant to be of low ability with probability one and I did not restrict ex-ante whether the less able, in fact, participate or not. However, for $\beta < 2$, the outside option turned out to dominate low effort for the less able and I reverted to two refinements to show robustness.

Suppose the evaluator's bias and performance uncertainty are sufficiently high to induce high-ability applicants to exert high effort and low-ability applicants to exert low effort. By Bayes' rule, if there is randomness in the environment, the evaluator believes the applicant to be of high ability with probability less than unity when observing a high-quality performance. As a result, the hiring probability at q_H is strictly smaller than in the fully separating equilibrium. This is intuitive: under partial separation, the evaluator is more cautious in his hiring policy when observing a high-quality performance due to the chance that the applicant behind the curtain is only of low ability. I define $\underline{\beta}_{sep}(p,\varepsilon) := \max\{0, \underline{\beta}_{sep'}(p,\varepsilon), \underline{\beta}_{sep''}(p,\varepsilon)\}$ where $\underline{\beta}_{sep'}(p,\varepsilon) := \{\beta : U(e_L,\eta_L) = 0\}$ and $\underline{\beta}_{sep''}(p,\varepsilon) := \{\beta : U(e_L,\eta_L) = U(e_H,\eta_L)\}$. Moreover, $\bar{\beta}_{sep}(p,\varepsilon) := \min\{2, \bar{\beta}_{sep'}(p,\varepsilon)\}$ where $\bar{\beta}_{sep'}(p,\varepsilon) := \{\beta : U(e_L,\eta_H) = U(e_H,\eta_H)\}$.

Lemma 4′. *Suppose Assumption 5 holds and $\varepsilon > \hat{\varepsilon}$. For $\beta \in \beta_B^H(\varepsilon) := (\underline{\beta}_{sep}(p,\varepsilon), \bar{\beta}_{sep}(p,\varepsilon)]$ where $\underline{\beta}_{sep}(p,\varepsilon) < \bar{\beta}_{sep}(p,\varepsilon)$, a partially separating equilibrium exists in which high-ability applicants exert high effort and low-ability applicants exert low effort.*

Proof of Lemma 4′. First, by Bayes' rule, when observing a high-quality performance, the evaluator believes the applicant to be of high ability with probability

$$\Pr(\eta_H|q_H) = \frac{p}{p+\varepsilon(1-p)} := \gamma,$$

and of low ability with complementary probability. At q_H, the evaluator, therefore, expects gross utility

$$\begin{aligned}
\mathbb{E}[V|q_H] = q_H + \mathbb{E}[\eta|q_H] - \frac{\beta}{2} &= 2 + \left[\frac{\varepsilon(1-p)}{p+\varepsilon(1-p)} + \frac{2p}{p+\varepsilon(1-p)}\right] - \frac{\beta}{2}, \\
&= 2 + \left[1 + \frac{p}{p+\varepsilon(1-p)}\right] - \frac{\beta}{2}, \\
&= 3 + \gamma - \frac{\beta}{2}
\end{aligned}$$

from hiring the applicant. Thus, the hiring probability at q_H modifies to

$$\Pr(h|q_H) = \Pr(\bar{U} \leqslant 3 + \gamma - \frac{\beta}{2}) = \frac{2(1+\gamma)+\beta}{4+2\beta}.$$

When taking into account the performance uncertainty in the mapping from low effort to performance quality, the expected utilities are

$$\begin{aligned}
U(e_L,\eta_L) &= \left[\frac{\beta}{4+2\beta}(1-\varepsilon) + \frac{2(1+\gamma)+\beta}{4+2\beta}\varepsilon\right]w - \frac{1}{2}, \\
U(e_H,\eta_L) &= \frac{2(1+\gamma)+\beta}{4+2\beta}w - 2, \\
U(e_L,\eta_H) &= \left[\frac{\beta}{4+2\beta}(1-\varepsilon) + \frac{2(1+\gamma)+\beta}{4+2\beta}\varepsilon\right]w - \frac{1}{4}, \\
U(e_H,\eta_H) &= \frac{2(1+\gamma)+\beta}{4+2\beta}w - 1,
\end{aligned}$$

in a partially separating blind audition, where $w := \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. I now show that, for $\beta \in \beta_B^H(\varepsilon) := (\underline{\beta}_{sep}(p,\varepsilon), \bar{\beta}_{sep}(p,\varepsilon)]$ where $\underline{\beta}_{sep}(p,\varepsilon) < \bar{\beta}_{sep}(p,\varepsilon)$, a partially separating equilibrium exists.

First, note that low-ability applicants must find it optimal to exert low effort. This gives the lower bound $\underline{\beta}_{sep}(p,\varepsilon)$ for which a partially separating equilibrium can be supported. For low degrees of performance uncertainty, the lower bound is determined by the indifference condition

$$U(e_L, \eta_L) = 0.$$

Given that $\beta \in [0,2]$, $\varepsilon \in [0,\frac{1}{2})$ and $p \in (0,1)$, the above equation is uniquely solved by $\underline{\beta}_{sep'}(p,\varepsilon) := \frac{\sqrt{p^2\varepsilon^4+4p^2\varepsilon^3-20p^2\varepsilon^2+16p^2\varepsilon-2p\varepsilon^4-12p\varepsilon^3+24p\varepsilon^2+\varepsilon^4+8\varepsilon^3}-p\varepsilon^2+4p\varepsilon-2p+\varepsilon^2-2\varepsilon}{p\varepsilon-p-\varepsilon}$. For larger degrees of performance uncertainty, the lower bound is determined by the indifference condition

$$U(e_L, \eta_L) = U(e_H, \eta_L).$$

Given that $\beta \in [0,2]$, $\varepsilon \in [0,\frac{1}{2})$ and $p \in (0,1)$, the above equation is uniquely solved by $\underline{\beta}_{sep''}(p,\varepsilon) := \frac{6(p\varepsilon^2-2p\varepsilon+p-\varepsilon^2)}{p\varepsilon^2-6p\varepsilon+5p-\varepsilon^2+4\varepsilon}$. Therefore, the lower bound of the partially separating equilibrium is given by $\underline{\beta}_{sep}(p,\varepsilon) := \max\{0, \underline{\beta}_{sep'}(p,\varepsilon), \underline{\beta}_{sep''}(p,\varepsilon)\}$.

Second, note that high-ability applicants must find it optimal to exert high effort. This gives the upper bound $\bar{\beta}_{sep}(p,\varepsilon)$ for which a partially separating equilibrium can be supported. For high degrees of performance uncertainty, the upper bound is determined by the indifference condition

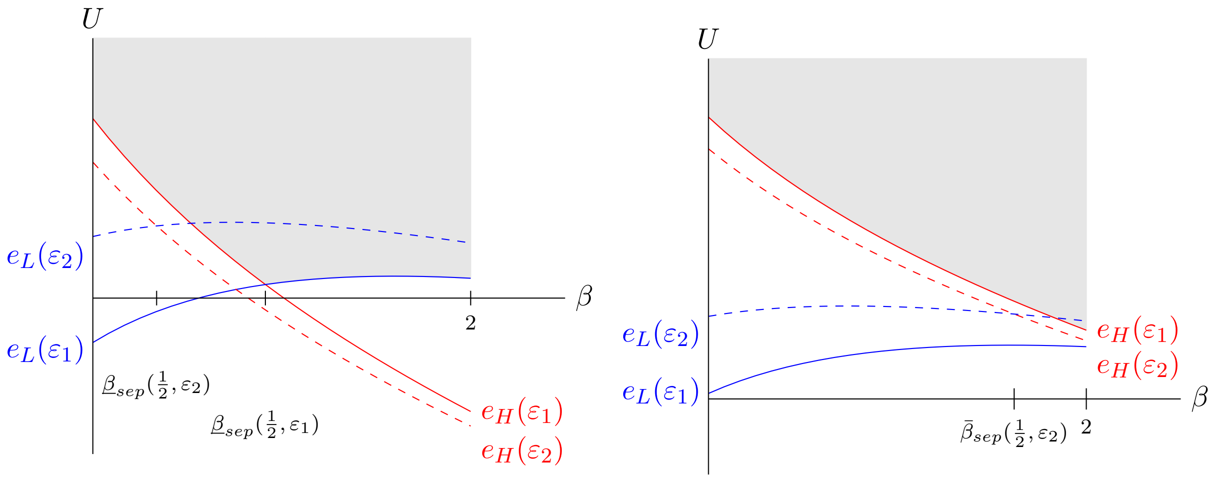
$$U(e_L, \eta_H) = U(e_H, \eta_H).$$

Given that $\beta \in [0,2]$, $\varepsilon \in [0,\frac{1}{2})$ and $p \in (0,1)$, the above equation is uniquely solved by $\bar{\beta}_{sep'}(p,\varepsilon) := \frac{6(2p\varepsilon^2-5p\varepsilon+3p-2\varepsilon^2+\varepsilon)}{2p\varepsilon^2-9p\varepsilon+7p-2\varepsilon^2+5\varepsilon}$. Therefore, the upper bound of the partially separating equilibrium is given by $\bar{\beta}_{sep}(p,\varepsilon) := \min\{2, \bar{\beta}_{sep'}(p,\varepsilon)\}$. □

Lemma 4′ implies that, for a fixed prior, once a certain degree of performance uncertainty is reached, partial separation is an equilibrium for a range of biases.[2] As the prior enters the applicant's expected utility via the conditional probability that she is of high ability given a high-quality performance, it is instructive to fix $p = \frac{1}{2}$ (Figure A.6). An increase in performance uncertainty from ε_1 to $\varepsilon_2 := \frac{3}{10}$ makes exerting low effort relatively more attractive for both high- and low-ability applicants. As a result, it reduces the lower bound of the partially separating equilibrium via the effort responses of the less able. Beyond a critical level of approximately 0.19, an increase in performance uncertainty also reduces the

[2] At $\varepsilon_0 = 0$, low-ability applicants are indifferent between exerting low effort and not participating if the evaluator's bias is maximal, $\beta = 2$. Thus, I defined $\beta_B^H := (\sqrt{17} - 3, 2)$ in Section 2.4 to support the fully separating equilibrium and neglected the possibility of partial separation at $\beta = 2$.

upper bound of the partially separating equilibrium via the effort responses of the highly able.



(a) Low-ability applicant. (b) High-ability applicant.

Figure A.6: Effort choices in a partially separating blind audition under performance uncertainty.

For the evaluator's belief that high-ability applicants exert high effort and low-ability applicants exert low effort to be consistent with the applicants' effort decision, he needs to have a sufficiently high bias, $\beta \in \beta_B^H(\varepsilon)$, and there needs to be randomness $\varepsilon > \hat{\varepsilon}$ in the environment.

Proposition 2′. *Suppose Assumption 5 holds, $\varepsilon > \hat{\varepsilon}$ and $\beta \in \beta_B^H(\varepsilon)$ where $\beta_B^H(\varepsilon)$ is defined as in Lemma 4′. Then, $\mathbb{E}[\Pi_B|\beta \in \beta_B^H(\varepsilon)] \gtrless 0$.*

Proof of Proposition 2′. The proof follows the steps detailed in the proof of Proposition 3 when there is no performance uncertainty. The evaluator's expected net utility is then obtained from simplifying

$$\begin{aligned}
&\mathbb{E}[\Pi_B|\beta \in \beta_B^H(\varepsilon)] = \\
&\quad (1-p)\left[\frac{\beta}{4+2\beta}(1-\varepsilon)\left(2-\frac{\beta}{2}-\left(3-\frac{\beta}{2}\right)\right)+\frac{2(1+\gamma)+\beta}{4+2\beta}\varepsilon\left(3-\frac{\beta}{2}-\left(3-\frac{\beta}{2}\right)\right)\right] \\
&\qquad\qquad + p\left[\frac{2(1+\gamma)+\beta}{4+2\beta}\left(4-\frac{\beta}{2}-\left(3-\frac{\beta}{2}\right)\right)\right], \\
&\qquad\qquad\qquad = (1-p)\left[-\frac{\beta}{4+2\beta}(1-\varepsilon)\right]+p\left[\frac{2(1+\gamma)+\beta}{4+2\beta}\right] \gtrless 0.
\end{aligned}$$

□

Proposition 2′ highlights that the evaluator's expected net utility in a partially separating equilibrium need not be positive.

A.1.3 Agency Cost

Definition 2.

$$AC_I(\beta,p) := \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right)+\left(1+\frac{\beta}{2}\right)\right] - \mathbb{E}[\Pi_I|\beta \in \beta_I^k] \quad \text{for} \quad k = L, M, H, \quad \text{and}$$

$$AC_B(\beta,p) := p\frac{4+\beta}{4+2\beta} - \mathbb{E}[\Pi_B|\beta \in \beta_B^j] \quad \text{for} \quad j = L, H.$$

Proposition 2. *Suppose Assumption 4 holds. Then, $AC_I(\beta,p) = 0$ if $\beta \in \beta_I^L$ and $AC_I(\beta,p) > 0$ if $\beta \in \beta_I^M \cup \beta_I^H$. Moreover, $AC_B(\beta,p) = 0$ if $\beta \in \beta_B^H$ and $AC_B(\beta,p) > 0$ if $\beta \in \beta_B^L$.*

Proof of Proposition 2. Under Assumption 4, the agency cost in an informed audition is calculated by substituting the evaluator's expected net utility for a low, moderate and high bias, detailed in Proposition 1, into Definition 2:

$$AC_I(\beta,p) = \begin{cases} 0 & \text{if } \beta \in \beta_I^L, \\ \frac{1-p}{2}\frac{\beta}{2+\beta}\left(1-\frac{\beta}{2}\right) & \text{if } \beta \in \beta_I^M, \\ \frac{1-p}{2}\frac{\beta}{2+\beta}\left(1-\frac{\beta}{2}\right)+\frac{p}{2}\left(\frac{4+\beta}{2+\beta}\right) & \text{if } \beta \in \beta_I^H. \end{cases}$$

The agency cost in a blind audition is calculated by substituting the evaluator's expected net utility for a low bias, detailed in Proposition 2, and for a high bias, detailed in Proposition 3, into Definition 2:

$$AC_B(\beta,p) = \begin{cases} 0 & \text{if } \beta \in \beta_B^H, \\ \frac{2p(1-p)}{4+2\beta} & \text{if } \beta \in \beta_B^L. \end{cases}$$

□

A.1.4 Gender-blind Curriculum Vitae

Given a gender-blind CV, low-ability applicants cannot pool with the highly able in the blind audition. Thus, I can consider the effort decision of low- and high-ability applicants separately.

Proposition 3. *For a low bias, $\beta \in \beta_G^L := [0, \frac{6}{5}]$, only high-ability applicants participate and exert high effort. The evaluator's expected net utility is $\mathbb{E}[\Pi_G|\beta \in \beta_G^L] = \mathbb{E}[\Pi_B|\beta \in \beta_B^H] > 0$.*

For a high bias, $\beta \in \beta_G^H := (\frac{6}{5}, 2)$, only high-ability applicants participate and exert low effort. The evaluator's expected net utility is $\mathbb{E}[\Pi_G|\beta \in \beta_G^H] = 0$.

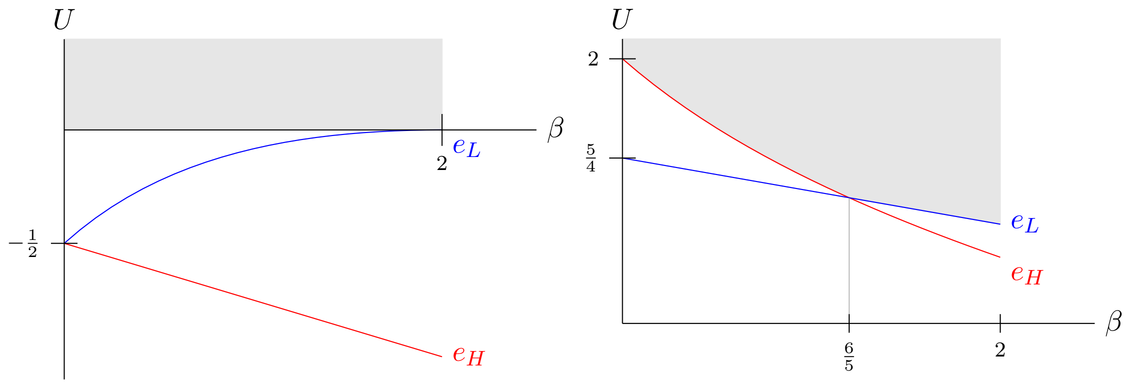
Proof of Proposition 3. With gender-blind CVs, hiring rule (2.7) modifies to $\Pr(\text{hired}|q,\eta) := \Pr(\bar{U} \leqslant \mathbb{E}[V|q,\eta]) = \frac{q+\eta+\frac{\beta}{2}-(2-\beta)}{2+\beta}$. Given Assumption 4 and the modified hiring rule, there are four possible hiring probabilities in a blind audition with gender-blind CVs:

$$\Pr(h|q_L,\eta_L) = \frac{\beta}{4+2\beta}, \quad \Pr(h|q_H,\eta_L) = \Pr(h|q_L,\eta_H) = \frac{2+\beta}{4+2\beta}, \quad \Pr(h|q_H,\eta_H) = \frac{4+\beta}{4+2\beta}.$$

Substituting for the four hiring probabilities in equation (2.2) yield

$$U(e_L, \eta_L) = \frac{\beta}{4+2\beta} w - \frac{1}{2}, \qquad U(e_H, \eta_L) = \frac{2+\beta}{4+2\beta} w - 2,$$
$$U(e_L, \eta_H) = \frac{2+\beta}{4+2\beta} w - \frac{1}{4}, \qquad U(e_H, \eta_H) = \frac{4+\beta}{4+2\beta} w - 1,$$

for both female and male applicants in a blind audition with gender-blind CVs, where $w := \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$. Low-ability applicants drop out of the audition for all $\beta < 2$ (Figure A.7a). High-ability applicants exert high effort if $\beta \in [0, \frac{6}{5}]$, and low effort if $\beta \in (\frac{6}{5}, 2]$ (Figure A.7b).



(a) Low-ability applicant. (b) High-ability applicant.

Figure A.7: Effort choices in a blind audition coupled with gender-blind CVs.

The calculation of the evaluator's expected net utility for $\beta \in \beta_G^L := [0, \frac{6}{5}]$ is identical to the calculation in the proof of Proposition 3. Therefore, $\mathbb{E}[\Pi_G | \beta \in \beta_G^L] = \mathbb{E}[\Pi_B | \beta \in \beta_B^H] > 0$. When the evaluator's bias against female applicants is high, $\beta \in \beta_G^H := (\frac{6}{5}, 2)$, the evaluator's expected net utility is $\mathbb{E}[\Pi_G | \beta \in \beta_G^H] := \mathbb{E}[V | \beta \in \beta_G^H] - w = p \Pr(h | q_L, \eta_H) \big[\mathbb{E}[V | q_L, \eta_H] - w\big] = 0$. □

A.2 Chapter 2 Additional Materials

A.2.1 Further Analysis in the Duopoly

A.2.1.1 Other Equilibria

In this section, I discuss two more equilibria: specialisation with a two-sided price discount and Bertrand competition in m. Together with a one-sided price discount, I show that the equilibria partition the parameter space of conscious shoppers' concern. Let $\bar{\theta} := \Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)$.

Proposition 4. *There exists an equilibrium in which F_i and F_j sell m to the whole market at $p_m = c(x_m^*)$ if $\theta < \bar{\theta}$. In the m-duopoloy, the firms exert optimal effort and make zero profit.*

There exists an equilibrium in which F_i and F_j specialise if $\theta \geqslant \bar{\theta}$. Conscious shoppers are offered a price discount on h. Quality shoppers may be offered a price discount on m.

Proof of Proposition 4. The proof proceeds in two steps. I first derive the conditions under which specialisation can be supported as an equilibrium when the price adjustment is assumed to be two- rather than one-sided. Together with the specialised equilibrium derived in Proposition 3 of the main text, this gives the cutoff $\bar{\theta}$. I then show that a duopoly in m is an equilibrium for $\theta < \bar{\theta}$.

First, suppose the price adjustment is two-sided in the specialised equilibrium. As before, the two firms must make the same profit. Let F_i offer discount ε_h on h to C, and F_j offer discount ε_m on m to Q. Q do not have an incentive to deviate to buying h if the discount ε_h is sufficiently low:

$$\begin{aligned} V_Q(p_h^{**}, x_h^{**}) + \varepsilon_h = V_Q(p_h^{**} - \varepsilon_h, x_h^{**}) &\leqslant V_Q(p_m^* - \varepsilon_m, x_m^*) = V_Q(p_m^*, x_m^*) + \varepsilon_m, \\ A_h x_h^{**} - (A_h x_h^{**} + \theta) + \varepsilon_h &\leqslant \varepsilon_m, \\ \varepsilon_h &\leqslant \varepsilon_m + \theta. \end{aligned} \tag{A.5}$$

C do not have an incentive to deviate to buying m if the discount ε_m is sufficiently low:

$$\begin{aligned} V_C(p_m^*, x_m^*) + \varepsilon_m = V_C(p_m^* - \varepsilon_m, x_m^*) &\leqslant V_C(p_h^{**} - \varepsilon_h, x_h^{**}) = V_C(p_h^{**}, x_h^{**}) + \varepsilon_h, \\ \varepsilon_m &\leqslant \varepsilon_h. \end{aligned} \tag{A.6}$$

Moreover, F_i cannot have an incentive to deviate and lower the price further to $p_h^* - \varepsilon_m$ such that Q switch to buying h:

$$\begin{aligned} \Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) = Q\Pi(p_m^* - \varepsilon_m, x_m^*; m) &\geqslant \Pi(p_h^* - \varepsilon_m, x_h^*; h), \\ Q\Pi(p_m^*, x_m^*; m) - Q\varepsilon_m &\geqslant \Pi(p_h^*, x_h^*; h) - \varepsilon_m, \\ \varepsilon_m &\geqslant \frac{1}{C}\Pi(p_h^*, x_h^*; h) - \frac{1-C}{C}\Pi(p_m^*, x_m^*; m). \end{aligned} \tag{A.7}$$

Also F_j cannot have an incentive to deviate and lower the price further to $p_m^* - \varepsilon_h$ such that C switch to buying m:

$$\begin{aligned} Q\Pi(p_m^* - \varepsilon_m, x_m^*; m) = \Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) &\geqslant \Pi(p_m^* - \varepsilon_h, x_m^*; m), \\ C\Pi(p_h^*, x_h^*; h) + C\theta - C\varepsilon_h &\geqslant \Pi(p_m^*, x_m^*; m) - \varepsilon_h, \\ \varepsilon_h &\geqslant \frac{1}{1-C}\Pi(p_m^*, x_m^*; m) - \frac{C}{1-C}[\Pi(p_h^*, x_h^*; h) + \theta]. \end{aligned} \tag{A.8}$$

The discounts equating profits satisfy $\Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) = Q\Pi(p_m^* - \varepsilon_m, x_m^*; m)$; that is,

$$C\Pi(p_h^*, x_h^*; h) + C\theta - C\varepsilon_h = Q\Pi(p_m^*, x_m^*; m) - Q\varepsilon_m. \tag{A.9}$$

Case I Suppose F_i's incentive compatibility constraint (A.7) binds. Substituting ε_m from (A.7) into the equal profit constraint (A.9) yields $\varepsilon_h = \left[\frac{1}{C} + \left(\frac{1-C}{C}\right)^2\right]\Pi(p_h^*, x_h^*; h) + \theta - \frac{1-C}{C^2}\Pi(p_m^*, x_m^*; m)$. The discounts need to be nonnegative for shoppers to buy. In particular, $\varepsilon_m \geqslant 0$ requires $\Pi(p_h^*, x_h^*; h) \geqslant (1-C)\Pi(p_m^*, x_m^*; m)$. The incentive compatibility constraint of C then ensures that $e_h \geqslant 0$ as well. I also need to check that the proposed discounts satisfy the other three incentive compatibility constraints. Substituting for ε_h and ε_m in (A.5) yields $\Pi(p_h^*, x_h^*; h) \leqslant \Pi(p_m^*, x_m^*; m)$, which is always satisfied given $c_m < c_h$. Substituting for ε_h and ε_m in (A.6) yields $\theta \geqslant \left(\frac{1-C}{C}\right)^2[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)]$. Substituting for ε_h and ε_m in (A.8) yields

$$\theta \geqslant \left[1 + \left(\frac{1-C}{C}\right)^2\right][\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)] =: \hat{\theta}.$$

This second lower bound on θ is more stringent. Thus, for $\Pi(p_h^*, x_h^*; h) \geqslant (1-C)\Pi(p_m^*, x_m^*; m)$, ε_h and ε_m are equilibrium discounts if $\theta \geqslant \hat{\theta}$. Each firm makes profit

$$\Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) = Q\Pi(p_m^* - \varepsilon_m, x_m^*; m) = \frac{1-C}{C}[\Pi(p_h^*, x_h^*; h) - \Pi(p_m^*, x_m^*; m)].$$

Case II Suppose F_j's incentive compatibility constraint (A.8) binds. Substituting ε_h from (A.8) into the equal profit constraint (A.9) yields $\varepsilon_m = \Pi(p_m^*, x_m^*; m) + \frac{C}{(1-C)^2}[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) - \theta]$. The discounts are nonnegative whenever

$$\theta \leqslant \left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) =: \check{\theta}.$$

I again need to check that the proposed discounts satisfy the other incentive compatibility constraints. Substituting for ε_h and ε_m in (A.5) yields $C^2[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)] \geqslant (2C-1)\theta$, which is always satisfied if $C \leqslant 1/2$. If $C > 1/2$, the constraint rearranges to $\theta \leqslant \frac{C^2}{2C-1}[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)]$. Substituting for ε_h and ε_m in (A.6) yields

$$\theta \geqslant \Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) =: \bar{\theta}.$$

Lastly, substituting for ε_h and ε_m in (A.7) yields

$$\theta \leqslant \left[1 + \left(\frac{1-C}{C}\right)^2\right][\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)] =: \hat{\theta}.$$

Note that $[1 + ((1 - C)/C)^2] \leqslant C^2/(2C - 1)$ whenever $C > 1/2$. Thus, $\hat{\theta}$ and $\check{\theta}$ apply as upper bounds on θ. In particular, for $\Pi(p_h^*, x_h^*; h) \geqslant (1 - C)\Pi(p_m^*, x_m^*; m)$, ε_h and ε_m are equilibrium discounts if $\bar{\theta} \leqslant \theta \leqslant \hat{\theta}$. For $\Pi(p_h^*, x_h^*; h) < (1 - C)\Pi(p_m^*, x_m^*; m)$, ε_h and ε_m are equilibrium discounts if $\bar{\theta} \leqslant \theta \leqslant \check{\theta}$. Each firm makes profit

$$\Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) = Q\Pi(p_m^* - \varepsilon_m, x_m^*; m) = \frac{C}{1 - C}[\Pi(p_h^*, x_h^*; h) + \theta - \Pi(p_m^*, x_m^*; m)].$$

In sum, if $\theta \geqslant \bar{\theta}$, specialisation can be supported as an equilibrium. For $\Pi(p_h^*, x_h^*; h) \geqslant (1 - C)\Pi(p_m^*, x_m^*; m)$, the price adjustment is always two-sided: **Case I** if $\theta \geqslant \hat{\theta}$, **Case II** if $\theta \leqslant \hat{\theta}$. For $\Pi(p_h^*, x_h^*; h) < (1 - C)\Pi(p_m^*, x_m^*; m)$, the specialised equilibrium with a one-sided price adjustment exists instead of Case I, and its lower bound equals the upper bound $\check{\theta}$ from **Case II**.

No Specialisation If $\theta < \bar{\theta}$, the incentive compatibility constraint for C fails: they switch to buying m, inducing F_i to produce m also. In trying to capture shoppers, the firms compete prices down to the common effort cost $c(x_m^*)$. In so doing, the firms keep effort at its optimum: for any effort $x'_m < x_m^*$, the competitor could instead exert x_m^* and offer its product at $p_m = c(x_m^*) + \epsilon$, where $\epsilon < \Pi(p_m^*, x_m^*; m) - [A_m x'_m - c(x'_m)]$. Obtaining a greater utility, shoppers would only buy the optimal-quality product, and the competitor would make profit $\epsilon > 0$.

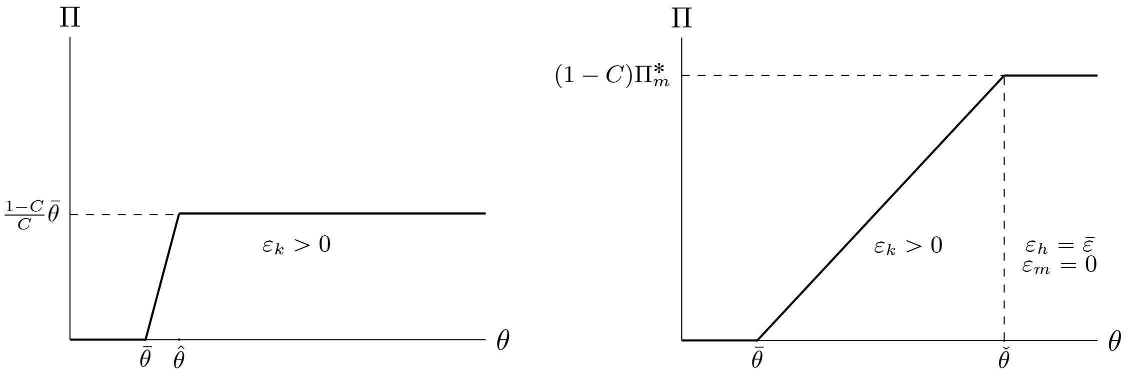
At $p_m = c(x_m^*)$ and $x_m^* = A_m/c_m$, both firms make zero profit but neither firm has an incentive to deviate to producing h. To make a strictly positive profit $C\epsilon$ from this deviation, a firm would need to offer h to C at price $p_h = c(x_h^*) + \epsilon$ with effort $x_h^* = A_h/c_h$. For $\theta < \bar{\theta}$, however, conscious shoppers do not switch back to buying h because their utility from buying m is greater:

$$A_h x_h^* + \theta - (c(x_h^*) + \epsilon) = V_C(c(x_h^*) + \epsilon, x_h^*) < V_C(c(x_m^*), x_m^*) = A_m x_m^* - c(x_m^*),$$
$$\Pi(p_h^*, x_h^*; h) + \theta - \epsilon < \Pi(p_m^*, x_m^*; m).$$

Hence, for $\theta < \bar{\theta}$, a duopoly in m with $\Pi(c(x_m^*), x_m^*; m) = 0$ can be supported as an equilibrium. □

Specialisation can be supported as an equilibrium for sufficiently high concerns. For lower concerns, both firms offer a machine-made product and compete prices down to the common effort cost. A firm has no incentive to produce by hand when the competitor exerts effort x_m^*: for any price p_h slightly above effort cost $c(x_h^*)$, conscious shoppers would not buy the handmade product.

For sufficiently high concerns, the firms can relax price competition through specialisation and thereby make a strictly positive profit. The handmade product has to be offered at a discount. Otherwise, the firm producing m has an incentive to switch the production process. The machine-made product may be sold at full price p_m^*. The firms' profits are capped, and only increase in conscious shoppers' concern up to a certain level (Figure A.8).



(a) $\Pi_h^* \geqslant (1-C)\Pi_m^*$. (b) $\Pi_h^* < (1-C)\Pi_m^*$.

Figure A.8: Equilibrium profit as a function of conscious shoppers' concern.

A.2.1.2 Sequential Moves in the Duopoly

In my benchmark model, I assume that the two firms decide simultaneously on the production process, pricing strategy and per-unit effort. In some settings, however, the production process may be flexible only in the long-run while effort and prices can be adjusted in the short-run. I model this timing as a two-stage game between the firms: in the first stage, $k \in \{h, m\}$ is chosen simultaneously. In the second stage, p_k and x_k are chosen simultaneously. In the second stage, each firm is aware whether its competitor produces by hand or machine. As before, shoppers have unit demand.

Proposition 5 (Specialised Equilibrium)**.** *There exists a subgame-perfect equilibrium in which F_i sells h to conscious shoppers at $p_h^{**} - \check{\varepsilon}$, and F_j sells m to quality shoppers at p_m^* if $\theta > (1-C)/C\Pi(p_h^*, x_h^*; h) + C\Pi(p_m^*, x_m^*; m)$. The price discount offered on h is $\check{\varepsilon} = C\Pi(p_m^*, x_m^*; m)$.*

Proof of Proposition 5. I solve this game by backward induction. In the second stage, each firm chooses p_k and x_k conditional on its own and its competitor's production process. I need to consider two classes of subgames. In the first class, F_i and F_j use identical production processes. Prices are competed down to the common effort cost $c(x_k^*)$. As before, effort x_k^* ensures that the competitor has no incentive to deviate. At $p_k = c(x_k^*)$, each firm makes zero profit.

In the second class, the firms use different production processes. Without loss of generality, suppose F_i announced $k = h$ and F_j announced $k = m$ in the first stage. Given that the firms cannot switch k, the firms' profits do not need to be equal. I consider three equilibrium candidates in this class: specialisation with a one- or a two-sided price reduction, and a monopoly in m.

One-sided Discount Let F_i offer discount $\breve{\varepsilon}$ on h. As before, C do not have an incentive to deviate to buying m. Q do not have an incentive to deviate to buying h if $\breve{\varepsilon} \leqslant \theta$. F_j does not have an incentive to deviate and lower the price to $p_m^* - \breve{\varepsilon}$ such that C switch to buying m if $\breve{\varepsilon} \geqslant C\Pi(p_m^*, x_m^*; m)$. F_i does not have an incentive to deviate and lower the price further to p_h^* such that Q switch to buying h if

$$\begin{aligned}\Pi(p_h^{**} - \breve{\varepsilon}, x_h^{**}; h) = C[\Pi(p_h^*, x_h^*; h) + \theta - \breve{\varepsilon}] \geqslant \Pi(p_h^*, x_h^*; h),\\ \theta \geqslant \frac{1-C}{C}\Pi(p_h^*, x_h^*; h) + \breve{\varepsilon}. \qquad \text{(A.10)}\end{aligned}$$

Now, let F_i make F_j's incentive compatibility constraint bind: $\breve{\varepsilon} = C\Pi(p_m^*, x_m^*; m) > 0$ leaves F_j better off setting its full price p_m^* and sell to Q only, rather than offering a price below $p_m^* - \breve{\varepsilon}$ to capture C also. Finally, I check that $\breve{\varepsilon}$ satisfies the other constraints. Given that F_i's incentive compatibility constraint is more stringent than Q's, substituting $\breve{\varepsilon}$ into (A.10) gives the lower bound $\tilde{\theta} := \frac{1-C}{C}\Pi(p_h^*, x_h^*; h) + C\Pi(p_m^*, x_m^*; m)$. If $\theta \geqslant \tilde{\theta}$, $\breve{\varepsilon}$ is an equilibrium discount. F_j makes profit $Q\Pi(p_m^*, x_m^*; m)$. F_i makes profit $\Pi(p_h^{**} - \breve{\varepsilon}, x_h^{**}; h) = C[\Pi(p_h^*, x_h^*; h) + \theta - C\Pi(p_m^*, x_m^*; m)]$.

Two-sided Discount Let F_i offer discount ε_h on h to C, and F_j offer discount ε_m on m to Q. Q do not have an incentive to deviate to buying h if $\varepsilon_h \leqslant \varepsilon_m + \theta$. C do not have an incentive to deviate to buying m if $\varepsilon_m \leqslant \varepsilon_h$. F_j does not have an incentive to deviate and lower the price further to $p_m^* - \varepsilon_h$ such that C switch to buying m if $\varepsilon_h \geqslant C\Pi(p_m^*, x_m^*; m) + Q\varepsilon_m$. F_i does not have an incentive to deviate and lower the price further to $p_h^* - \varepsilon_m$ such that Q switch to buying h if $\varepsilon_m \geqslant (1-C)\Pi(p_h^*, x_h^*; h) - C\theta + C\varepsilon_h$.

Now, let the firms' incentive compatibility constraints bind. Solve the two equations for discounts

$$\begin{aligned}\varepsilon_h &= \frac{1}{C + (1-C)^2}[C\Pi(p_m^*, x_m^*; m) + (1-C)^2\Pi(p_h^*, x_h^*; h) - C(1-C)\theta],\\ \varepsilon_m &= \frac{1}{C + (1-C)^2}[C^2\Pi(p_m^*, x_m^*; m) + (1-C)\Pi(p_h^*, x_h^*; h) - C\theta].\end{aligned}$$

As before, the discounts are nonnegative whenever $\varepsilon_m \geqslant 0$; that is, if $\theta \leqslant \tilde{\theta}$. Finally, I check that the proposed discounts satisfy the other incentive compatibility constraints. Substituting for ε_h and ε_m in $\varepsilon_m \leqslant \varepsilon_h$ yields $\frac{1-C}{C}[\Pi(p_h^*, x_h^*; h) - \Pi(p_m^*, x_m^*; m)] \leqslant \theta$ which is always satisfied given $c_m < c_h$. Substitution into $\varepsilon_h \leqslant \varepsilon_m + \theta$ yields $\theta \geqslant C\bar{\theta}$ where

$\bar{\theta} := [\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)]$. Thus, if $C\bar{\theta} \leqslant \theta \leqslant \tilde{\theta}$, ε_h and ε_m are equilibrium discounts with profits

$$\Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h) = \frac{C}{C + (1-C)^2}[\theta - C\bar{\theta}] \geqslant 0,$$
$$Q\Pi(p_m^* - \varepsilon_m, x_m^*; m) = \frac{1-C}{C + (1-C)^2}[C\theta + (1-C)\bar{\theta}] > 0.$$

No Specialisation If $\theta < C\bar{\theta}$, the incentive compatibility constraint for Q fails: they switch to buying h. This triggers price competition: F_j lowers its price to such an extent that F_i cannot even induce C to buy h at $p_h = c(x_h^*)$ under effort x_h^* as their utility from buying m is greater:

$$A_h x_h^* + \theta - c(x_h^*) = V_C(c(x_h^*), x_h^*) < V_C(p_m, x_m^*) = A_m x_m^* - p_m,$$
$$p_m < \frac{A_m^2}{c_m} - [\Pi(p_h^*, x_h^*; h) + \theta] =: p_M.$$

Specifically, F_j offers a slightly lower price than p_M, and makes profit

$$\Pi(p_M - \epsilon, x_m^*; m) = \Pi(p_m^*, x_m^*; m) - [\Pi(p_h^*, x_h^*; h) + \theta + \epsilon] = \bar{\theta} - [\theta + \epsilon]. \quad \text{(A.11)}$$

F_j does not have an incentive to deviate, and sell to Q only given $p_h = c(x_h^*)$ and x_h^*. In particular, F_j can set at most price $p_m = A_m^2/c_m - \Pi(p_h^*, x_h^*; h)$ to induce Q to buy m, and make profit

$$\Pi(A_m^2/c_m - \Pi(p_h^*, x_h^*; h), x_m^*; m) = (1-C)[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)] = (1-C)\bar{\theta}. \quad \text{(A.12)}$$

However, this deviation is not profitable: (A.11) is at least as great as (A.12) if $\theta + \epsilon \leqslant C\bar{\theta}$. Hence, a monopoly in m is an equilibrium for $\theta < C\bar{\theta}$. F_i does not produce, and makes zero profit.

In the first stage, the firms compare their maximised profits across the subgames (Example A.1). If conscious shoppers' concern is high or moderate, there are two pure-strategy subgame-perfect equilibria in which the firms choose different production processes. Both firms produce but do not necessarily make the same profit. If conscious shoppers' concern is low, there are three pure-strategy equilibria. Either one or two firms produce. Only m is sold to the whole market.

		F_i	
		h	m
F_j	h	$0, 0$	$\underline{A}, \underline{B}$
	m	$\underline{B}, \underline{A}$	$0, 0$

(a) High: $\tilde{\theta} \leqslant \theta$,
$A := \Pi(p_h^{**} - \check{\varepsilon}, x_h^{**}; h)$,
$B := Q\Pi(p_m^{*}, x_m^{*}; m)$.

		F_i	
		h	m
F_j	h	$0, 0$	$\underline{C}, \underline{D}$
	m	$\underline{D}, \underline{C}$	$0, 0$

(b) Moderate: $C\bar{\theta} \leqslant \theta \leqslant \tilde{\theta}$,
$C := \Pi(p_h^{**} - \varepsilon_h, x_h^{**}; h)$,
$D := Q\Pi(p_m^{*} - \varepsilon_m, x_m^{*}; m)$.

		F_i	
		h	m
F_j	h	$0, 0$	$\underline{0}, \underline{E}$
	m	$\underline{E}, \underline{0}$	$\underline{0}, \underline{0}$

(c) Low: $\theta < C\bar{\theta}$,
$E := \Pi(p_M - \epsilon, x_m^{*}; m)$.

Table A.1: The firms' maximised profits across subgames for different levels of concern.

□

Proposition 5 highlights that the firms can soften price competition whenever the conscious shoppers' concern θ is sufficiently high. In this case, the firms act as monopolists in their respective markets and cater to different segments. The firms, however, do not necessarily make the same profit: an equal profit constraint is only required if the production process is as flexible as the effort level and the price. This one additional constraint in the benchmark model implies that the one-sided discount $\bar{\varepsilon}$ is at least as large as the one-sided discount $\check{\varepsilon}$ offered on the handmade product under the alternative timing. Now, the firm offering h sees its profit rise when conscious shoppers become more concerned about production conditions: the discount $\check{\varepsilon}$ is independent of θ.

Therefore, my result on specialisation is qualitatively robust to a situation in which a firm is aware whether its competitor is engaging in hand production.

A.2.2 Further Analysis in the Signalling Game

A.2.2.1 Other pure-strategy equilibria

In this section, I analyse all other possible pure-strategy equilibria. First, note that there is no separating equilibrium in which the optimistic firm sells a machine-made product to the whole market, and the pessimistic firm sells a premium-priced handmade product to steadfastly conscious shoppers. Given the conformists' updated belief, they act like quality shoppers upon observing h, and act like conscious shoppers upon observing m. Consequently, the optimistic firm that knows the fraction of conscious shoppers among the steadfast to be high does not want to deviate to selling a handmade product if

$$\Pi(p_m^{*}, x_m^{*}; m) \geqslant \varphi_H(1 - \eta)\big[\Pi(p_h^{*}, x_h^{*}; h) + \theta\big] \quad \Leftrightarrow \quad \eta \geqslant \bar{\eta}(\varphi_H), \tag{A.13}$$

where

$$\bar{\eta}(\varphi) := 1 - \frac{1}{\varphi}\frac{\Pi(p_m^{*}, x_m^{*}; m)}{\Pi(p_h^{*}, x_h^{*}; h) + \theta}.$$

The pessimistic firm that knows the fraction of conscious shoppers among the steadfast to be low does not want to deviate to selling a machine-made product if

$$\varphi_L(1-\eta)\big[\Pi(p_h^*, x_h^*; h)+\theta\big] \geqslant \Pi(p_m^*, x_m^*; m) \quad \Leftrightarrow \quad \eta \leqslant \bar{\eta}(\varphi_L). \tag{A.14}$$

Given that $\bar{\eta}(\varphi_L) < \bar{\eta}(\varphi_H)$, condition (A.13) and condition (A.14) cannot be satisfied simultaneously.

In any pooling equilibrium, conformists cannot infer the firm's private information. Their behaviour, thus, depends on $\mathbb{E}(\varphi) = \gamma\varphi_H + (1-\gamma)\varphi_L \lesseqgtr 1/2$ with commonly known prior $\gamma := \Pr(\varphi = \varphi_H) \in (0,1)$. In particular, when conformists expect the median steadfast shopper to be conscious, they act like conscious shoppers in any pooling equilibrium. Conversely, when they do not expect the median steadfast shopper to be conscious, they act like quality shoppers.

Proposition 6 (Pooling Equilibrium M)**.** *If $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_H - \Pi(p_h^*, x_h^*; h)$, a pooling equilibrium exists in which the firm sells a machine-made product to the whole market whatever its private information if and only if $\eta \leqslant \hat{\eta}(\varphi_H)$, where $\hat{\eta}(\varphi_H) \in (0,1)$.*

Proof of Proposition 6. The conjectured pooling equilibrium exists if and only if the firm prefers m whatever its private information. Given Assumption 2, the firm's equilibrium profit $\Pi(p_m^*, x_m^*; m) = \frac{A_m^2}{2c_m}$ does not depend on conformists' prior belief. However, their off-path belief upon observing a deviation to $k = h$ determines the size of the parameter space for which pooling on machine production can be supported as an equilibrium. In what follows, I argue that it is a reasonable restriction for conformists to believe such deviation to come from a firm that privately observes φ_H.

First, the firm observing φ_H will be strictly better off deviating from hand to machine production if conformists in response to $k = h$ choose to act like conscious shoppers and buy a premium-priced handmade product with probability

$$\varphi_H(1-\eta)\Big[\Pi(p_h^*, x_h^*; h)+\theta\Big] + y\eta\Big[\Pi(p_h^*, x_h^*; h)+\theta\Big] > \Pi(p_m^*, x_m^*; m),$$
$$y > \frac{1}{\eta}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi_H(1-\eta)\right] := A.$$

Moreover, the firm observing φ_L will be indifferent or strictly better off deviating from hand to machine production if conformists choose to act like conscious shoppers and buy a premium-priced handmade product with probability $y \geqslant \frac{1}{\eta}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi_L(1-\eta)\right] := B$. Since $\varphi_H > 1/2 > \varphi_L$, $A < B$ which, in turn, implies $(A, 1] \supset [B, 1]$. Consequently, applying Divinity (Banks and Sobel, 1987), any belief that puts some positive probability on the pessimistic firm after observing a deviation to $k = h$ has to be discarded.[3]

Given the refined off-path belief $\Pr(\varphi_H|h) = 1$, the firm prefers machine production whatever its private information if

$$\Pi(p_m^*, x_m^*; m) \geqslant [\varphi(1-\eta)+\eta]\big[\Pi(p_h^*, x_h^*; h)+\theta\big] \quad \Leftrightarrow \quad \eta \leqslant \hat{\eta}(\varphi),$$

where

$$\hat{\eta}(\varphi) := \frac{1}{1-\varphi}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi\right].$$

Given that $\hat{\eta}(\varphi_H) < \hat{\eta}(\varphi_L)$, pooling is an equilibrium for $\eta \leqslant \min\{\hat{\eta}(\varphi_H), \hat{\eta}(\varphi_L)\} = \hat{\eta}(\varphi_H)$. Finally, note that $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_H - \Pi(p_h^*, x_h^*; h)$ ensures $\hat{\eta}(\varphi_H) > 0$ while Assumption 3 ensures $\hat{\eta}(\varphi_H) < 1$. □

Proposition 7 (Pooling Equilibrium H)**.** *Suppose $\mathbb{E}(\varphi) > 1/2$. A pooling equilibrium exists in which the firm sells a premium-priced handmade product to steadfastly conscious and conformist shoppers whatever its private information if and only if $\eta \geqslant \max\{0, \hat{\eta}(\varphi_L)\}$, where $\hat{\eta}(\varphi_L) < 1$.*

Suppose $\mathbb{E}(\varphi) < 1/2$. If $\theta > \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$, a pooling equilibrium exists in which the firm sells a premium-priced handmade product to steadfastly conscious shoppers whatever its private information if and only if $\eta \leqslant \bar{\eta}(\varphi_L)$.

Proof of Proposition 7. The conjectured equilibrium exists if and only if the firm prefers production by hand whatever its private information. Given that h can never be optimal for a firm selling to the whole market, I assume that $\eta \geqslant \tilde{\eta}(\varphi_L)$ if $\mathbb{E}(\varphi) > 1/2$, and $\eta \leqslant \dot{\eta}(\varphi_L)$ if $\mathbb{E}(\varphi) < 1/2$, where

$$\tilde{\eta}(\varphi_L) := \frac{1}{1-\varphi_L}\left[\frac{\Pi(p_h^*, x_h^*; h)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi_L\right] \quad \text{and} \quad \dot{\eta}(\varphi_L) := 1 - \frac{1}{\varphi_L}\frac{\Pi(p_h^*, x_h^*; h)}{\Pi(p_h^*, x_h^*; h)+\theta},$$

implying that $p_h^{**} = A_h x_h^* + \theta$ and $x_h^{**} = x_h^* = A_h/c_h$ are optimal.

First, suppose $\mathbb{E}(\varphi) > 1/2$. Then, conformists act like conscious shoppers in the pooling equilibrium, and are willing to pay a premium for the handmade product. Thus, the firm's equilibrium profit is $\Pi(p_h^{**}, x_h^{**}; h) = [\varphi(1-\eta)+\eta]\Big[\frac{A_h^2}{2c_h}+\theta\Big]$. Given Assumption 2, the off-path belief upon observing a deviation to $k = m$ is unrestricted. Intuitively, given a one-sided concern, conscious and quality shoppers' maximum willingness to pay for the

[3] I use Divinity rather than the Intuitive Criterion (Cho and Kreps, 1987) because h is not equilibrium-dominated for the φ_L-firm whenever $\eta \geqslant \hat{\eta}(\varphi_L)$. Therefore, the Intuitive Criterion is not strong enough to refine the off-path belief, and discard pooling on m for $\eta \geqslant \max\{\hat{\eta}_L, \bar{\eta}_L\}$ under prior belief $\mathbb{E}(\varphi) < 1/2$. A handmade strategy, however, is more costly for the φ_L-firm than the φ_H-firm so that it is more plausible that the latter deviates to producing h.

machine-made product coincides. Therefore, it does not matter what conformists believe off-path: their behaviour upon observing a deviation to $k = m$ is independent of their off-path belief. Hence, the firm prefers production by hand whatever its private information if

$$[\varphi(1-\eta)+\eta]\big[\Pi(p_h^*, x_h^*; h)+\theta\big] \geqslant \Pi(p_m^*, x_m^*; m) \quad \Leftrightarrow \quad \eta \geqslant \hat{\eta}(\varphi).$$

Given that $\hat{\eta}(\varphi_H) < \hat{\eta}(\varphi_L)$, pooling is an equilibrium for $\eta \geqslant \max\{0, \hat{\eta}(\varphi_H), \hat{\eta}(\varphi_L)\} = \max\{0, \hat{\eta}(\varphi_L)\}$.

Finally, I show that my assumption $\eta \geqslant \tilde{\eta}(\varphi_L)$ is implied by $\eta \geqslant \max\{0, \hat{\eta}(\varphi_L)\}$ so that neither the pessimistic nor the optimistic firm prefers deviating to p_h^* to sell also to quality shoppers:

$$\Pi(p_h^*, x_h^*; h) < \Pi(p_m^*, x_m^*; m),$$
$$\tilde{\eta}(\varphi_L) := \frac{1}{1-\varphi_L}\left[\frac{\Pi(p_h^*, x_h^*; h)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi_L\right] < \frac{1}{1-\varphi_L}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\theta} - \varphi_L\right] =: \hat{\eta}(\varphi_L).$$

Now, suppose $\mathbb{E}(\varphi) < 1/2$. Then, conformists act like quality shoppers in the pooling equilibrium, and are not willing to pay a premium for the handmade product. Thus, the firm's equilibrium profit is $\Pi(p_h^{**}, x_h^{**}; h) = \varphi(1-\eta)\left[\frac{A_h^2}{2c_h}+\theta\right]$. As before, the off-path belief is unrestricted. Hence, the firm prefers production by hand whatever its private information if

$$\varphi(1-\eta)\big[\Pi(p_h^*, x_h^*; h)+\theta\big] \geqslant \Pi(p_m^*, x_m^*; m) \quad \Leftrightarrow \quad \eta \leqslant \bar{\eta}(\varphi).$$

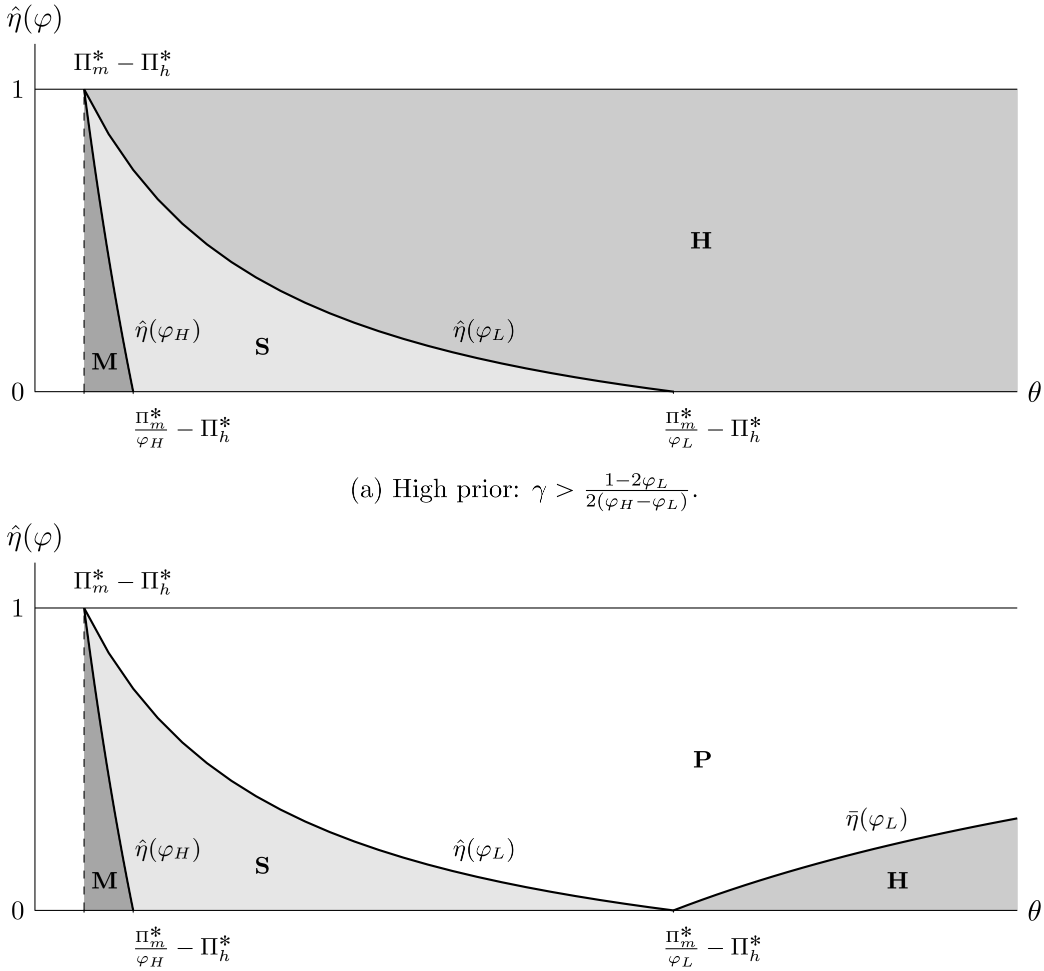
Given that $\bar{\eta}(\varphi_L) < \bar{\eta}(\varphi_H)$, pooling is an equilibrium for $\eta \leqslant \min\{\bar{\eta}(\varphi_H), \bar{\eta}(\varphi_L)\} = \bar{\eta}(\varphi_L)$. Note that $\theta > \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$ ensures $\bar{\eta}(\varphi_L) > 0$. Moreover, $\frac{d\bar{\eta}(\varphi_L)}{d\theta} > 0$ and $\lim_{\theta\to\infty} \bar{\eta}(\varphi_L) = 1$ together ensure $\bar{\eta}(\varphi_L) < 1$.

Finally, I show that my assumption $\eta \leqslant \dot{\eta}(\varphi_L)$ is implied by $\eta \leqslant \bar{\eta}(\varphi_L)$ so that neither the pessimistic nor the optimistic firm prefers deviating to p_h^* to sell also to quality shoppers and conformists:

$$\Pi(p_h^*, x_h^*; h) < \Pi(p_m^*, x_m^*; m),$$
$$\bar{\eta}(\varphi_L) := 1 - \frac{1}{\varphi_L}\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\theta} < 1 - \frac{1}{\varphi_L}\frac{\Pi(p_h^*, x_h^*; h)}{\Pi(p_h^*, x_h^*; h)+\theta} =: \dot{\eta}(\varphi_L).$$

□

The pooling equilibrium with production by hand depends on whether conformists expect the median steadfast shopper to be conscious: this belief determines whether the firm can charge conformists a premium for the handmade product. Given conscious shoppers' one-sided concern, pooling on machine production, in contrast, does not depend on the conformists' prior belief.



(a) High prior: $\gamma > \frac{1-2\varphi_L}{2(\varphi_H-\varphi_L)}$.

(b) Low prior: $\gamma < \frac{1-2\varphi_L}{2(\varphi_H-\varphi_L)}$.

Figure A.9: Equilibria in the signalling game for $\varphi_H = 0.8$ and $\varphi_L = 0.2$.

Given the refinement applied to the off-path belief under pooling on machine production, the three equilibria do not overlap (Figure A.9). Thus, the separating equilibrium is the unique pure-strategy equilibrium in the parameter space S. Moreover, when conformists expect the median steadfast shopper to be conscious, the three equilibria partition the parameter space (Figure A.9a). For a low prior, in contrast, no pure-strategy equilibrium exists for $\eta > \max\{\hat{\eta}(\varphi_L), \bar{\eta}(\varphi_L)\}$ (Figure A.9b).

Proposition 8 (Partially Separating Equilibrium P)**.** *Suppose* $\mathbb{E}(\varphi) < 1/2$. *A partially separating equilibrium exists in which the firm sells a premium-priced handmade product with probability one after observing* φ_H*, and with probability*

$$x = \frac{\gamma}{1-\gamma}\frac{2\varphi_H - 1}{1 - 2\varphi_L} \in (0, 1)$$

after observing φ_L *if and only if* $\eta \geqslant \max\{\hat{\eta}(\varphi_L), \bar{\eta}(\varphi_L)\}$*, where* $\max\{\hat{\eta}(\varphi_L), \bar{\eta}(\varphi_L)\} \in [0,1)$*. Conformists update their belief to* $\mathbb{E}(\varphi|h) = 1/2$ *and* $\mathbb{E}(\varphi|m) = \varphi_L$*. Upon observing a handmade strategy, conformists act like conscious shoppers with probability*

$$y = \frac{1}{\eta}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h) + \theta} - \varphi_L(1-\eta)\right] \in [0,1].$$

Proof of Proposition 8. Suppose $\mathbb{E}(\varphi) < 1/2$, and let the firm privately observing φ_H choose $k = h$ with certainty. First, I derive the probability x with which the firm privately observing φ_L can choose $k = h$ to induce conformists to update their belief to one-half. In particular, I require

$$\mathbb{E}(\varphi|h) = \Pr(\varphi_H|h)\varphi_H + (1 - \Pr(\varphi_H|h))\varphi_L = \varphi_L + \Pr(\varphi_H|h)(\varphi_H - \varphi_L) = \frac{1}{2} \tag{A.15}$$

for conformists to be indifferent between acting like conscious or quality shoppers. By Bayes' rule,

$$\Pr(\varphi_H|h) = \frac{\gamma}{\gamma + (1-\gamma)x}. \tag{A.16}$$

Substituting for the updated belief (A.16) in equation (A.15) yields the probability $x = \frac{\gamma}{1-\gamma}\frac{2\varphi_H - 1}{1-2\varphi_L}$.

Second, I derive the probability y with which the conformists act like conscious shoppers to make the firm privately observing φ_L indifferent between machine production and production by hand:

$$\Pi(p_m^*, x_m^*; m) = \varphi_L(1-\eta)\left[\Pi(p_h^*, x_h^*; h) + \theta\right] + y\eta\left[\Pi(p_h^*, x_h^*; h) + \theta\right],$$

$$y = \frac{1}{\eta}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h) + \theta} - \varphi_L(1-\eta)\right].$$

Lastly, the requirement of x and y being a valid probabilities yields the following parameter restrictions:

$$x > 0 \Leftrightarrow \varphi_H > 1/2 > \varphi_L \text{ and } \gamma \in (0,1), \qquad y \geqslant 0 \Leftrightarrow \eta \geqslant \bar{\eta}(\varphi_L),$$

$$x < 1 \Leftrightarrow \mathbb{E}(\varphi) < 1/2, \qquad y \leqslant 1 \Leftrightarrow \eta \geqslant \hat{\eta}(\varphi_L).$$

Thus, I require also $\eta \geqslant \max\{\hat{\eta}(\varphi_L), \bar{\eta}(\varphi_L)\}$ for the partially separating equilibrium to exist. □

Proposition 8 highlights that partial separation is an equilibrium in the parameter space P in Figure A.9b for which no pure-strategy equilibrium exists. Intuitively, when the conformists' prior belief is low, the pessimistic firm that knows the fraction of conscious shoppers among the steadfast to be low cannot always choose a handmade strategy to induce conformists to act like conscious shoppers. Instead, it uses the maximum probability x that induces conformists to update their belief to exactly one-half upon observing production by hand.

At this point, conformists are indifferent between acting like conscious or quality shoppers, and consequently they randomise between the two options. The probability y with which conformists act like conscious shoppers, in turn, makes the firm privately observing φ_L indifferent between the two production processes. Hence, this pessimistic firm is willing to randomise also. The higher the prior γ in the partially separating equilibrium, the higher the probability with which the pessimistic firm chooses a handmade strategy.

A.2.2.2 Signalling Benchmark: No private Information.

When the fraction of conscious shoppers among the steadfast is observable, there is no room for signalling. Conformists know whether the fraction is high or low when making the buying decision. Given the assumption that $\varphi_H > 1/2 > \varphi_L$, conformists act like conscious shoppers when knowing that most steadfast shoppers are conscious, and like quality shoppers otherwise.

Proposition 9. *If $\varphi \in \{\varphi_H, \varphi_L\}$ is observable, separation is an equilibrium for a greater proportion of the parameter space.*

Proof of Proposition 9. First, consider the pessimistic firm. Given that φ_L is observable, this firm decides between selling m to all or h to conscious shoppers only. It chooses machine production if

$$\Pi(p_m^*, x_m^*; m) \geqslant \varphi_L(1-\eta)\Big[\Pi(p_h^*, x_h^*; h) + \theta\Big] \Leftrightarrow \eta \geqslant \bar{\eta}(\varphi_L).$$

Now, consider the optimistic firm. Given that φ_H is observable, this firm decides between selling m to all or h to conscious and conformist shoppers. Therefore, this firm chooses hand production if

$$[\varphi_L(1-\eta) + \eta]\Big[\Pi(p_h^*, x_h^*; h) + \theta\Big] \geqslant \Pi(p_m^*, x_m^*; m) \Leftrightarrow \eta \geqslant \hat{\eta}(\varphi_H).$$

Thus, when there is no private information, separation is an equilibrium for $\eta \geqslant \max\{0, \hat{\eta}(\varphi_H), \bar{\eta}(\varphi_L)\}$. In the signalling game, in contrast, if $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$, separation is an equilibrium for $\eta \in [\max\{0, \hat{\eta}(\varphi_H)\}, \hat{\eta}(\varphi_L)]$.

Finally, to show that the parameter space of the separating equilibrium is larger without private information, note that $\max\{0, \hat{\eta}(\varphi_H), \bar{\eta}(\varphi_L)\} = \max\{0, \hat{\eta}(\varphi_H)\}$ if $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$. Consequently, if $\theta < \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$, the lower bound on η when φ is observable coincides with the lower bound when φ is private information. The upper bound $\hat{\eta}(\varphi_L)$ in the signalling game, however, is strictly below unity. □

Proposition 9 implies that the pessimistic firm pools more often with the optimistic firm on a handmade strategy if it has private information. In particular, the parameter space S above the dashed line in Figure A.10 supports either a pooling equilibrium on h for a high prior (Figure A.9a), or a partially separating equilibrium for a low prior (Figure A.9b) in the signalling game.

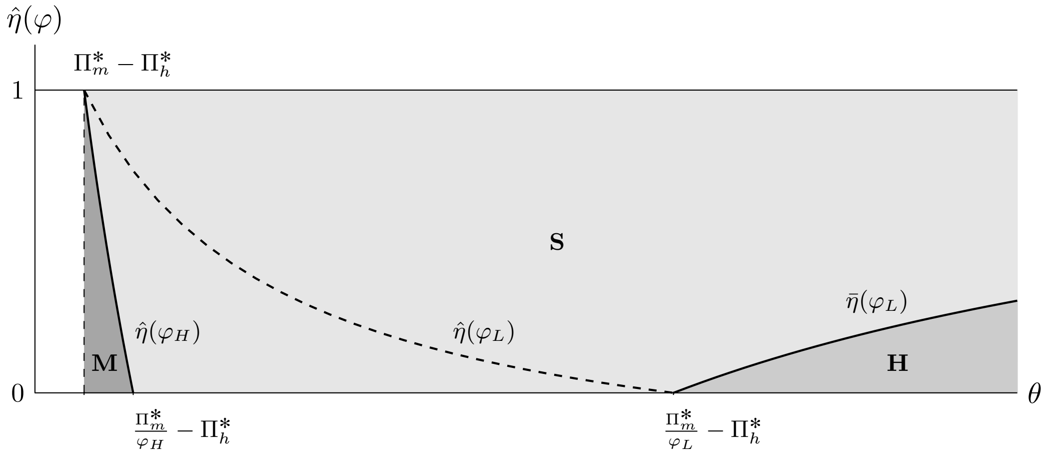


Figure A.10: Equilibria in the benchmark for $\varphi_H = 0.8$ and $\varphi_L = 0.2$.

A.2.2.3 Conformists' Utility Function

In this section, I show that my main insights extend to an alternative specification of the conformists' utility function that implies their willingness to pay to increase smoothly in the expected fraction of conscious shoppers among the steadfast. In particular, I assume that the conformists' utility function is a convex combination of the quality and conscious shoppers' utility function,

$$V_{CON}(a;k) = a[\beta V_C + (1-\beta)V_Q] = a[q(x_k,k) - p_k + \beta\theta\nu(k)],$$

where $a \in \{0,1\}$ is the conformists' buying decision and $\beta := (\mathbb{E}(\varphi|k) - \varphi_L)/(\varphi_H - \varphi_L)$ is the weight placed on conscious shoppers. Intuitively, the conformists' utility from buying a handmade product lies in between the utility gained by quality and conscious shoppers. If conformists expect the fraction of conscious shoppers among the steadfast to be low upon observing a handmade strategy, $\mathbb{E}(\varphi|h) = \varphi_L$, their willingness to pay is as low as quality shoppers' willingness to pay. Conversely, if they expect the fraction to be high, $\mathbb{E}(\varphi|h) = \varphi_H$, they are willing to pay as much as conscious shoppers. Moreover, their willingness to pay is now increasing linearly in $\mathbb{E}(\varphi|h)$; that is,

$$p_{CON,h} := \begin{cases} A_h x_h + \theta & if \;\; \mathbb{E}(\varphi|h) > 1/2, \\ A_h x_h & if \;\; \mathbb{E}(\varphi|h) < 1/2, \end{cases} \quad \text{becomes} \quad p_{CON,h} := A_h x_h + \frac{\mathbb{E}(\varphi|h) - \varphi_L}{\varphi_H - \varphi_L}\theta.$$

In the separating equilibrium, conformists infer the fraction of conscious shoppers among the steadfast to be high upon observing a handmade strategy, and to be low upon observing machine production. Consequently, their willingness to pay in the separating equilibrium is unaffected by the change in the utility function. Moreover, given that quality and conscious shoppers' willingness to pay coincides for a machine-made product under Assumption 2, pooling on m is unaffected.

The specification, however, alters the pooling equilibrium that involves a handmade strategy: when the firms pool on h, they now have the choice whether to sell to conformists and conscious shoppers at an intermediate price $p'_h = A_h x^*_h + \gamma\theta$, or to conscious shoppers only at a higher price $p^{**}_h = A_h x^*_h + \theta$. Intuitively, when the optimistic and the pessimistic firm pool on a handmade strategy, conformists' willingness to pay falls short of conscious shoppers' willingness to pay. Thus, the firm has the option to charge a full handmade premium, or a handmade premium adjusted for how likely the firm is optimistic ex-ante, $\gamma := \Pr(\varphi = \varphi_H) \in (0, 1)$.

Proposition 10 (Pooling Equilibrium H′)**.** *If $\theta > (\Pi(p^*_m, x^*_m; m) - \Pi(p^*_h, x^*_h; h))/\gamma$, a pooling equilibrium exists in which the firm sells a handmade product at an intermediate price p'_h to steadfastly conscious and conformist shoppers whatever its private information if and only if $\eta \geqslant \max\{\check{\eta}(\varphi_L), \eta'(\varphi_H)\}$.*

If $\theta > \Pi(p^*_m, x^*_m; m)/\varphi_L - \Pi(p^*_h, x^*_h; h)$, a pooling equilibrium exists in which the firm sells a premium-priced handmade product to steadfastly conscious shoppers whatever its private information if and only if $\eta \leqslant \min\{\bar{\eta}(\varphi_L), \eta'(\varphi_L)\}$.

Proof of Proposition 10. The optimistic and the pessimistic firm must prefer the same pricing. In particular, the intermediate price $p'_h = A_h x^*_h + \gamma\theta$ is optimal if

$$[\varphi(1-\eta)+\eta]\Big[\Pi(p^*_h, x^*_h; h) + \gamma\theta\Big] \geqslant \varphi(1-\eta)\Big[\Pi(p^*_h, x^*_h; h) + \theta\Big],$$

$$\eta \geqslant \frac{\varphi(1-\gamma)\theta}{[\varphi(1-\gamma)+\gamma]\theta + \Pi(p^*_h, x^*_h; h)} =: \eta'(\varphi).$$

Given that this cutoff is increasing in the fraction of conscious shoppers among the steadfast,

$$\frac{d\eta'(\varphi)}{d\varphi} = \frac{(1-\gamma)\gamma\theta^2 + (1-\gamma)\theta\Pi(p^*_h, x^*_h; h)}{\big[[\varphi(1-\gamma)+\gamma]\theta + \Pi(p^*_h, x^*_h; h)\big]^2} > 0,$$

the optimistic and the pessimistic firm prefer the intermediate price p'_h if $\eta \geqslant \max\{\eta'(\varphi_H), \eta'(\varphi_L)\} = \eta'(\varphi_H)$. Conversely, both prefer the high price p^{**}_h if $\eta \leqslant \min\{\eta'(\varphi_H), \eta'(\varphi_L)\} = \eta'(\varphi_L)$. As before, Assumption 2 makes $\mathbb{E}(\varphi|k)$ irrelevant when observing a deviation to $k = m$. Hence, the off-path belief remains unrestricted in this alternative specification.

See the proof of Proposition 7 also for showing that the firm prefers losing out on quality shoppers.

First, suppose $\eta \geqslant \eta'(\varphi_H)$. The firm prefers hand production whatever its private information if

$$[\varphi(1-\eta)+\eta]\big[\Pi(p_h^*, x_h^*; h)+\gamma\theta\big] \geqslant \Pi(p_m^*, x_m^*; m) \quad \Leftrightarrow \quad \eta \geqslant \check{\eta}(\varphi),$$

where

$$\check{\eta}(\varphi) := \frac{1}{1-\varphi}\left[\frac{\Pi(p_m^*, x_m^*; m)}{\Pi(p_h^*, x_h^*; h)+\gamma\theta} - \varphi\right].$$

Given that $\check{\eta}(\varphi_H) < \check{\eta}(\varphi_L)$, pooling is an equilibrium for $\eta \geqslant \max\{\check{\eta}(\varphi_L), \eta'(\varphi_H)\}$. Note that $\theta > (\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h))/\gamma$ ensures $\check{\eta}(\varphi_L) < 1$. Moreover, Assumption 3 ensures $\eta'(\varphi_H) \in (0,1)$.

Now, suppose $\eta \leqslant \eta'(\varphi_L)$. The firm prefers hand production whatever its private information if

$$\varphi(1-\eta)\big[\Pi(p_h^*, x_h^*; h)+\theta\big] \geqslant \Pi(p_m^*, x_m^*; m) \quad \Leftrightarrow \quad \eta \leqslant \bar{\eta}(\varphi).$$

Given that $\bar{\eta}(\varphi_L) < \bar{\eta}(\varphi_H)$, pooling is an equilibrium for $\eta \leqslant \min\{\bar{\eta}(\varphi_L), \eta'(\varphi_L)\}$. As before, $\theta > \Pi(p_m^*, x_m^*; m)/\varphi_L - \Pi(p_h^*, x_h^*; h)$ ensures $\bar{\eta}(\varphi_L) > 0$. Assumption 3 ensures $\eta'(\varphi_L) \in (0,1)$. □

The key requirement is that the optimistic and the pessimistic firm make the same pricing decision for the handmade product: either the intermediate price p' or the high price p^{**} is optimal for both. Otherwise, the price becomes a credible signal and conformists are able to infer the firm's private information, adjusting their willingness to pay.

Proposition 10 highlights that, compared to $\mathbb{E}(\varphi) > 1/2$ under the initial specification, pooling on a handmade strategy is an equilibrium for a smaller proportion of the parameter space: the white area in Figure A.11 no longer supports pooling on h. Moreover, as the prior γ approaches unity, the partition depicted in Figure A.11 approaches the partition when the median steadfast shopper is conscious (Figure A.9a). The optimistic and the pessimistic firm pool more often on a handmade strategy with price p' sold to conscious and conformist shoppers: the area H_1' increases. The firms pool less often on a handmade strategy with price p^{**}: the area H_2' vanishes in the limit.

Intuitively, the conformists' willingness to pay, and thus price p', is increasing in the prior. Hence, a large-scale handmade strategy targeting conscious and conformist shoppers becomes more appealing than an exclusive handmade strategy targeting conscious shoppers only.

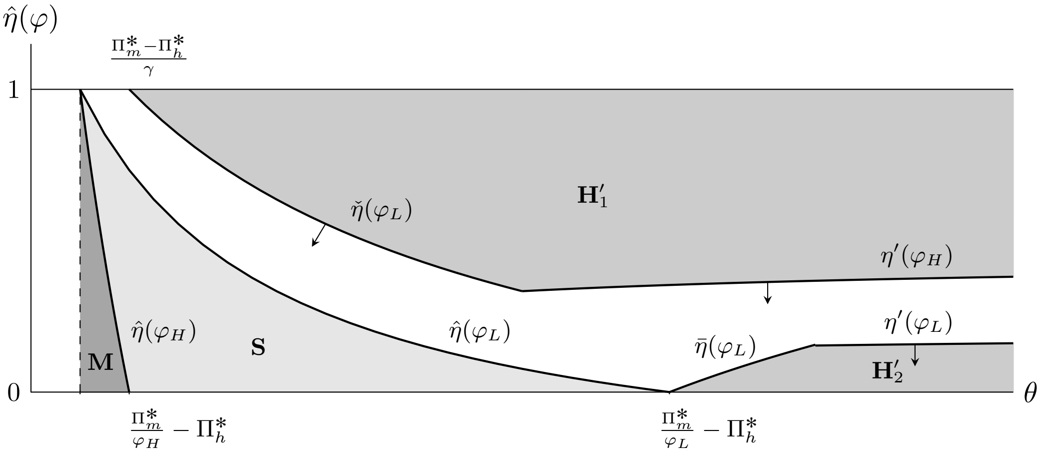


Figure A.11: Equilibria in the modified signalling game for $\varphi_H = 0.8$ and $\varphi_L = 0.2$.

A.3 Chapter 3 Additional Materials

A.3.1 Derivation of Thresholds

Below we will derive both $\lambda_i(M_i, \theta_i)$ and $\lambda_i^R(h_i, \theta_i)$. Player i's expected utility will be key in both derivations and is given by

$$\mathbb{E}[U(s_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(s_i - \bar{s}_{-i})^2 | h_i, s_i, \theta_i] - \lambda_i(s_i - \theta_i)^2.$$

Fixing player i's action as $s_i = \theta_i$ or $s_i = 1 - \theta_i$ gives

$$\mathbb{E}[U(\theta_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(\theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, \theta_i] \tag{A.17}$$

and

$$\mathbb{E}[U(1-\theta_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, 1-\theta_i] - \lambda_i(1-2\theta_i)^2,$$

respectively. For both $\theta_i = 0$ and $\theta_i = 1$, we have $(1-2\cdot 0)^2 = (1-2\cdot 1)^2 = 1$. Therefore, we write player i's expected utility when $s_i = 1-\theta_i$ as

$$\mathbb{E}[U(1-\theta_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | h_i, \theta_i, 1-\theta_i] - \lambda_i, \tag{A.18}$$

which includes the part $-\lambda_i(1-2\theta_i)^2 = -\lambda_i$.

We now provide the detailed steps required to derive $\lambda_i(M_i, \theta_i)$.

(i) When player i is credulous, she believes that all other players truthfully reveal their type: $s^*(\theta_j) = \theta_j$ because $\hat{\lambda}_{ij} = 1$ for all $j \neq i \in \mathcal{N}$. Therefore, we replace the history h_i with the summary statistic $M_i := \sum_{j \in h_i} \mathbb{I}\{s_j = \theta_i\}$.

(ii) Another implication of the belief that $s^*(\theta_j) = \theta_j$ is that we can remove the conditioning of $\mathbb{E}[(s_i - \bar{s}_{-i})^2 | M_i, \theta_i, s_i]$ on s_i, because, from player i's credulous perspective, her action s_i does not affect the other players' actions. This means that we can re-write (A.17) and (A.18) as

$$\mathbb{E}[U(\theta_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i]$$

and

$$\mathbb{E}[U(1-\theta_i; \theta_i, \lambda_i)] = -(1-\lambda_i)\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i] - \lambda_i.$$

(iii) We open up the brackets in $\mathbb{E}[(\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i]$ and $\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i]$ to obtain

$$\mathbb{E}[(\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i] = \mathbb{E}[\theta_i^2 - 2\theta_i \bar{s}_{-i} + \bar{s}_{-i}^2 | M_i, \theta_i] \tag{A.19}$$

and

$$\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i] = \mathbb{E}[(1-\theta_i)^2 - 2(1-\theta_i)\bar{s}_{-i} + \bar{s}_{-i}^2 | M_i, \theta_i], \tag{A.20}$$

respectively. Then, because the terms θ_i^2 and $(1-\theta_i)^2$ are held fixed, we can re-write (A.19) and (A.20) as

$$\mathbb{E}[(\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i] = \theta_i^2 + \mathbb{E}[\bar{s}_{-i}^2 - 2\theta_i \bar{s}_{-i} | M_i, \theta_i] \tag{A.21}$$

and

$$\mathbb{E}[(1-\theta_i - \bar{s}_{-i})^2 | M_i, \theta_i] = (1-\theta_i)^2 + \mathbb{E}[\bar{s}_{-i}^2 - 2(1-\theta_i)\bar{s}_{-i} | M_i, \theta_i], \tag{A.22}$$

respectively.

(iv) The term $\lambda_i(M_i, \theta_i)$ is defined by $\mathbb{E}[U(\theta_i; \theta_i, \lambda_i)] \geqslant \mathbb{E}[U(1-\theta_i; \theta_i, \lambda_i)]$. Thus, substituting (A.21) and (A.22) into (A.17) and (A.18) yields

$$-(1-\lambda_i)\big[\theta_i^2 + \mathbb{E}[\bar{s}_{-i}^2 - 2\theta_i \bar{s}_{-i} | M_i, \theta_i]\big] \geqslant -(1-\lambda_i)\big[(1-\theta_i)^2 + \mathbb{E}[\bar{s}_{-i}^2 - 2(1-\theta_i)\bar{s}_{-i} | M_i, \theta_i]\big].$$

Then, given step (3), we can cancel the term $-(1-\lambda_i)\mathbb{E}[\bar{s}_{-i}^2 | M_i, \theta_i]$ from both sides:

$$-(1-\lambda_i)\big[\theta_i^2 + \mathbb{E}[-2\theta_i \bar{s}_{-i} | M_i, \theta_i]\big] \geqslant -(1-\lambda_i)\big[(1-\theta_i)^2 + \mathbb{E}[-2(1-\theta_i)\bar{s}_{-i} | M_i, \theta_i]\big]. \tag{A.23}$$

Consequently, due to the linearity of the expectations operator, we can write (A.23) as

$$-(1-\lambda_i)\big[\theta_i^2 - 2\theta_i \mathbb{E}[\bar{s}_{-i} | M_i, \theta_i]\big] \geqslant -(1-\lambda_i)\big[(1-\theta_i)^2 - 2(1-\theta_i)\mathbb{E}[\bar{s}_{-i} | M_i, \theta_i]\big]. \tag{A.24}$$

(v) Finally, we rearrange (A.24) for λ_i:

$$\lambda_i \geqslant -(1-\lambda_i)[(1-\theta_i)^2-\theta_i^2]+(1-\lambda_i)2(1-2\theta_i)\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i],$$
$$\lambda_i \geqslant -(1-\lambda_i)(1-2\theta_i)+(1-\lambda_i)2(1-2\theta_i)\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i],$$
$$\lambda_i \geqslant (1-2\theta_i)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big)-\lambda_i(1-2\theta_i)\big(\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big). \tag{A.25}$$

To proceed, we divide (A.25) by λ_i's common factors: $1+(1-2\theta_i)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big)$. Therefore, we need to show that this term is positive to not reverse the inequality sign. First, note that $\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]\in[0,1]$ for $\theta_i\in\{0,1\}$. Consequently, we have

$$1+2\mathbb{E}[\bar{s}_{-i}|M_i,0]-1=2\mathbb{E}[\bar{s}_{-i}|M_i,0]\geqslant 0$$

and

$$1+(-1)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,1]-1\big)+1=2-2\mathbb{E}[\bar{s}_{-i}|M_i,1]\geqslant 0.$$

Thus, dividing by these common factors yields

$$\lambda_i \geqslant \frac{(1-2\theta_i)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big)}{1+(1-2\theta_i)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big)},$$
$$\lambda_i \geqslant 1-\frac{1}{1+(1-2\theta_i)\big(2\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]-1\big)},$$
$$\lambda_i \geqslant 1-\frac{1}{2\big(\theta_i+(1-2\theta_i)\mathbb{E}[\bar{s}_{-i}|M_i,\theta_i]\big)},$$

where the right-hand side of the inequality is defined as $\lambda_i(M_i,\theta_i)$ in Section 4.2.1.

We now provide the detailed steps to derive $\lambda_i^R(h_i,\theta_i)$.

(i) Since player i is rational, we cannot use steps (1)-(3) from the derivation of $\lambda_i(M_i,\theta_i)$. Instead, we compare (A.17) and (A.18) as they are:

$$-(1-\lambda_i)\mathbb{E}[(\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,\theta_i]\geqslant-(1-\lambda_i)\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i]-\lambda_i.$$

(ii) Expanding the brackets gives

$$\lambda_i+\lambda_i\mathbb{E}[(\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,\theta_i]-\lambda_i\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i]$$
$$\geqslant \mathbb{E}[(\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,\theta_i]-\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i].$$

(iii) As $\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i]\leqslant 1$, rearranging for λ_i gives

$$\lambda_i \geqslant \frac{\mathbb{E}[(\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,\theta_i]-\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i]}{1+\mathbb{E}[(\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,\theta_i]-\mathbb{E}[(1-\theta_i-\bar{s}_{-i})^2|h_i,\theta_i,1-\theta_i]},$$

where the right-hand side of the inequality is defined as $\lambda_i^R(h_i,\theta_i)$ in Section 4.2.1.

A.3.2 An Alternative Model of Scepticism

In this section, we set up a parameterised model that (i) places a sceptical player on a continuum between a rational and credulous player, (ii) allows for heterogeneity in the complexity of heuristics within a group, and (iii) relaxes the assumption that a player using a heuristic is wrong about other players' conformity motives with probability one.

In particular, we assume that all players have an interior conformity motive $\lambda \in (0, 1)$. A sceptical player i, however, presumes that there are not one but *two* kinds of players in the group: with probability ε_i, player $j \neq i$ is only concerned with matching her type ($\hat{\lambda}_{ij} = 1$); and, with probability $1-\varepsilon_i$, player $j \neq i$ is as concerned as player i about fitting in ($\hat{\lambda}_{ij} = \lambda$). The parameter ε_i is common knowledge.[4] It can be interpreted as the degree to which player i simplifies (4.1). As the simplification in the assessment of every other player's preferences approaches zero (one), a sceptical player i becomes rational (credulous). Consequently, the rational and credulous model are limit cases of the parameterised model when all (interior) players make correct assessments with probability one and zero, respectively.

For comparability to Section 4.4, we focus on a group of four players. We assume that player 2 and 3 are sceptical, and that $1 - \varepsilon_2$ and $1 - \varepsilon_3$ characterise the complexity of their heuristics. For player 4 it does not matter whether she is credulous, sceptical or rational about others' concern to fit in as she does not face any successors. For player 1, we assume that she follows $s_1^*(\theta_1) = \theta_1$, whether she is credulous, sceptical or rational.[5] Define

$$\lambda_2(\varepsilon_2, h_{20}) := \frac{5 - 2\varepsilon_2 - 2\pi(1-\pi)(1-\varepsilon_2)}{14 - 2\varepsilon_2 - 2\pi(1-\pi)(1-\varepsilon_2)} < \frac{9 - 2\pi(1-\pi)(5+\varepsilon_3)}{18 - 2\pi(1-\pi)(5+\varepsilon_3)} =: \lambda_3(\varepsilon_3, h_{30})$$

as the cut-offs of a sceptical player 2 at history $h_{20} = \{1 - \theta_2\}$ and a sceptical player 3 at history $h_{30} = \{1 - \theta_2, 1 - \theta_2\}$, respectively.

Theorem 1′. *The mapping between types and actions is:*

(i) $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *if* $\lambda \in \big[0, \lambda_2(\varepsilon_2, h_{20})\big)$.

(ii) $(\theta, \theta, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $(\theta, 1-\theta, \theta_3, \theta_4) \mapsto (\theta, 1-\theta, \theta_3, \theta_4)$ *if* $\lambda \in \big[\lambda_2(\varepsilon_2, h_{20}), \lambda_3(\varepsilon_3, h_{30})\big)$.

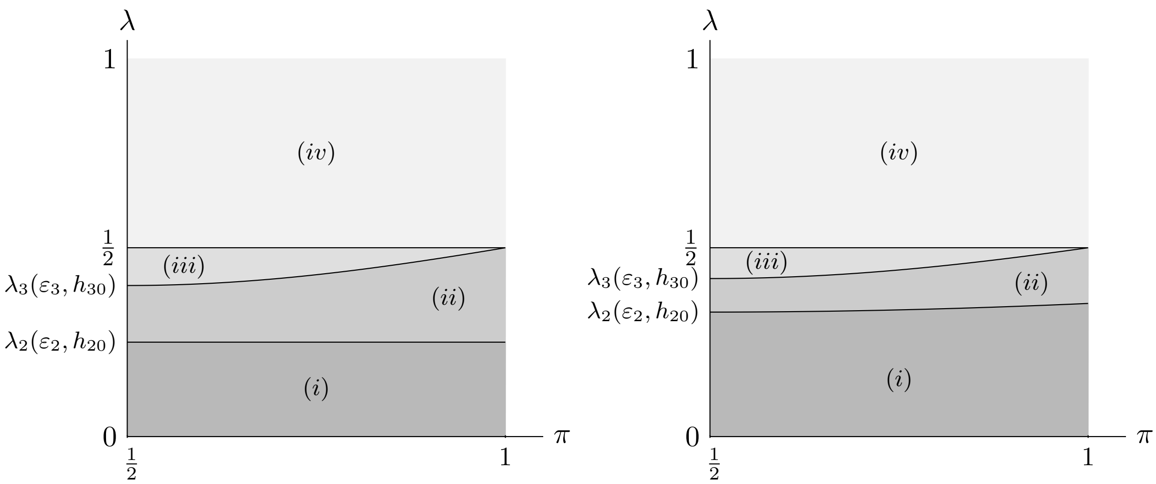
(iii) $(\theta, \theta, \theta, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $\{(\theta, \theta, 1-\theta, \theta_4), (\theta, 1-\theta, \theta, \theta_4), (\theta, 1-\theta, 1-\theta, \theta_4)\} \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if* $\lambda \in \big[\lambda_3(\varepsilon_3, h_{30}), 1/2\big)$.

[4] If, say, $\varepsilon_i = 0.3$ and $\varepsilon_j = 0.4$, player i presumes that player j is only concerned with matching her type with probability 0.3, and player i knows that player j's presumption about i's conformity motive is correct with probability 0.6.

[5] The calculation of player 1's expected utility is algebraically intractable given the 2^6 possible combinations of $(\theta_2, \theta_2, \theta_4; \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4)$; however, given that a sceptical player is a convex combination of a credulous and a rational player, who both truthfully reveal when moving first, the assumption that $s_1^*(\theta_1) = \theta_1$ is innocuous.

(iv) $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if* $\lambda \in \big[1/2, 1\big]$.

Compared to our definition of scepticism in the main text, now a sceptical player's cut-off for a given history always lies between her credulous and rational cut-off, confirming the intuition that credulity and rationality obtain as limit cases of scepticism in this alternative specification: $\lambda_2(0, \theta_2) \leqslant \lambda_2(\varepsilon_2, h_{20}) \leqslant \lambda_2^R(h_{20}, \theta_2)$ and $\lambda_3(0, \theta_3) \leqslant \lambda_3(\varepsilon_3, h_{30}) \leqslant \lambda_3^R(h_{20}, \theta_3)$ for all $\pi \in (1/2, 1)$. As a result, when player 2 and 3 change from being credulous to being rational about the other players' conformity motive, mappings (ii) and (iii) in Proposition 1′ cover less of the admissible parameter space (π, λ) whereas mapping (i) covers more. In particular, Figure A.12 shows that, when the players simplify less, the interval $\big[0, \lambda_2(\varepsilon_2, h_{20})\big)$ on which mapping (i) is the aggregate outcome of the interaction becomes larger whilst the intervals $\big[\lambda_3(\varepsilon_3, h_{30}), 1/2\big)$ of mapping (ii) and $\big[\lambda_3(\varepsilon_3, h_{30}), 1/2\big)$ of mapping (iii) become smaller for any given $\pi \in (1/2, 1)$.



(a) If credulous ($\varepsilon_2 = \varepsilon_3 = 1$). (b) If almost rational ($\varepsilon_2 = \varepsilon_3 = 0.05$).

Figure A.12: The aggregate outcome in a homogeneous group of four players.

We can, thus, conclude that whenever the true conformity motive λ has a distribution F with full support, a fall in ε_2 and ε_3 makes uniform behaviour ex-ante more likely. Intuitively, uniform behaviour is possible in mappings (ii) and (iii) but certain in mapping (i), and some probability mass shifts from the former to the latter as the heuristics used within the group become more complex. Put differently, as in Corollary 1, a rationality boost in the sense of a fall in ε_2 and ε_3 from one to zero makes uniform behaviour more likely. However, contrary to Corollary 1, a scepticism boost in the sense of a fall in ε_2 and ε_3 from one to $x \in (0, 1)$ also makes uniform behaviour more likely. In the main text, a scepticism boost reduced the probability of uniform behaviour.

Given the parameterisation of scepticism by ε_i, we can now accommodate heterogeneity in the degree to which players simplify (4.1). Making only player 2's heuristic more complex (ε_2 falls while keeping ε_3 fixed) reduces the area covered by mapping (*ii*) and increases the area covered by mapping (*i*). Making only player 3's heuristic more complex reduces the area covered by mapping (*iii*) and increases the area covered by mapping (*ii*). Hence, the above conclusions also hold for a targeted boost.

Proof of Theorem 1′. We solve backwards from player 4. For the last player, it is irrelevant whether she is credulous, sceptical or rational. At history $\{1-\theta_4, 1-\theta_4, 1-\theta_4\}$, she falsifies if $\lambda < 1/2$. At histories $\{1-\theta_4, 1-\theta_4, \theta_4\}$, $\{1-\theta_4, \theta_4, 1-\theta_4\}$ and $\{\theta_4, 1-\theta_4, 1-\theta_4\}$, she falsifies if $\lambda < 1/4$. Hence, for $\lambda \geqslant 1/2$ we can immediately conclude that the mapping is $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ as social pressure from uniform behaviour is (i) increasing over time and (ii) never high enough to at least affect the last player's behaviour.

Player 3 infers that player 1 is always truthful, and that with probability ε_3 player 2 is only concerned with matching her type whilst with probability $1-\varepsilon_3$ player 2 has the same conformity motive $\lambda \in (0,1)$.

Case I Suppose first that player 3 takes into account that λ is so low that she cannot infer player 2's type θ_2 when player 2 has $\lambda_2 = \lambda$. Then, at history $h_{30} = \{1-\theta_3, 1-\theta_3\}$, $\Pr(\omega = 1-\theta_3 | h_{30}, \theta_3) = \pi$ with probability ε_3; $\Pr(\omega = 1-\theta_3 | h_{30}, \theta_3) = 1/2$ with probability $1-\varepsilon_3$. If player 3 follows her predecessors, she infers that player 4 with $\lambda_4 = \lambda$ will do so, too. If she instead truthfully reveals, her action reduces the social pressure so that player 4 with $\lambda_4 = \lambda$ will not follow either. Hence, at $h_{30} = \{1-\theta_3, 1-\theta_3\}$, she expects

$$\mathbb{E}[U_3^S(\theta_3; \theta_3, \lambda)|h_{30}] = -\varepsilon_3\left[\pi^2 + (1-\pi)^2 + \pi(1-\pi)\frac{8}{9}\right](1-\lambda) - (1-\varepsilon_3)\left[\frac{13}{18}\right](1-\lambda)$$

from being truthful whereas

$$\mathbb{E}[U_3^S(1-\theta_3; \theta_3, \lambda)|h_{30}] = -\lambda - \varepsilon_3\left[\frac{2\pi(1-\pi)\varepsilon_3}{9}\right](1-\lambda) - (1-\varepsilon_3)\left[\frac{\varepsilon_3}{18}\right](1-\lambda)$$

from falsifying. Therefore, she falsifies if

$$\lambda < \frac{13 + \varepsilon_3[4 + \varepsilon_3 - 4\pi(1-\pi)(5+\varepsilon_3)]}{18 + \varepsilon_3[4 + \varepsilon_3 - 4\pi(1-\pi)(5+\varepsilon_3)]}.$$

When $\varepsilon_3 = 0$, the above cut-off is independent of π. The less complex the heuristic, the more steeply the cut-off is increasing in π: the impact of the complexity of the heuristic on the cut-off is small in moderately informative environments but significant when π is takes on an extreme value.

Given that we assume player 3 cannot infer player 2's type when $\lambda_2 = \lambda$, we now find the upper bound on λ such that $\mathbb{E}[U_2^S(1-\theta_2; \theta_2, \lambda)|h_{20}] > \mathbb{E}[U_2^S(\theta_2; \theta; \lambda)|h_{20}]$ at $h_{20} = \{1-\theta_2\}$.

Given that player 2 infers player 1 to be truthful, $\Pr(w = 1 - \theta_2 | h_{20}, \theta_2) = 1/2$. Player 2 needs to infer both player 3's and player 4's action. If player 2 follows her predecessor, she infers that player 3 with $\lambda_3 = \lambda$ will do so, too. If she instead truthfully reveals, her action reduces the social pressure so that player 3 with $\lambda_3 = \lambda$ will not follow either. What player 2 predicts player 4 to do depends on whether the true common conformity motive λ is above or below 1/4 (see above analysis).

(a) Assuming first that $\lambda \geqslant 1/4$, player 2 predicts player 4 with $\lambda_4 = \lambda$ to be truthful when she is truthful. Hence, at $h_{20} = \{1 - \theta_2\}$, she expects

$$\mathbb{E}[U_2^S(\theta_2; \theta_2, \lambda) | h_{20}] = -\left[\frac{5 - 2\pi(1-\pi)}{9}\right](1 - \lambda)$$

from being truthful whereas

$$\mathbb{E}[U_2^S(1 - \theta_2; \theta_2, \lambda) | h_{20}] = -\lambda - \varepsilon_2 \left[\frac{2(1 - \pi(1 - \pi))}{9}\right](1 - \lambda)$$

from falsifying. Therefore, she falsifies if $\lambda < \lambda_2(\varepsilon_2, h_{20})$, where

$$\lambda_2(\varepsilon_2, h_{20}) := \frac{5 - 2\varepsilon_2 - 2\pi(1-\pi)(1-\varepsilon_2)}{14 - 2\varepsilon_2 - 2\pi(1-\pi)(1-\varepsilon_2)} \geqslant \frac{1}{4}.$$

When $\varepsilon_2 = 1$, the cut-off is independent of π. The more complex the heuristic, the more steeply the cut-off is increasing in π. Hence, the impact of the complexity of the heuristic is least (most) pronounced in uninformative (informative) environments as $\pi \to 1/2$ (1). Given that $\lambda_2(\varepsilon_2, h_{20})$ is lower than player 3's cut-off, we conclude that the mapping is $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ if $\lambda \in \big[1/4, \lambda_2(\varepsilon_2, h_{20})\big)$.

(b) Assuming now that $\lambda < 1/4$, player 2 instead predicts player 4 with $\lambda_4 = \lambda$ to always agree with the majority. In this case, $\mathbb{E}[U_2^S(1 - \theta_2; \theta_2, \lambda) | h_{20}]$ is as above but

$$\mathbb{E}[U_2^S(\theta_2; \theta_2, \lambda) | h_{20}] = -(1 - \varepsilon_2)\left[\frac{5}{9}\right](1 - \lambda)$$

so that she falsifies if $\lambda < (5 - 2\varepsilon_2)/(14 - 2\varepsilon_2)$. Given that this cut-off is weakly greater than 1/4 for all ε_2, we conclude that the mapping is also $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ if $\lambda \in \big[0, 1/4\big)$.

Case II Now suppose that player 3 takes into account that λ is sufficiently high that she can infer player 2's type θ_2 when player 2 has $\lambda_2 = \lambda$. Then, at history $h_{30} = \{1 - \theta_3, 1 - \theta_3\}$, $\Pr(\omega = 1 - \theta_3 | h_{30}, \theta_3) = \pi$ with probability 1. Hence, she expects

$$\mathbb{E}[U_3^S(\theta_3; \theta_3, \lambda) | h_{30}] = -\left[\pi^2 + (1 - \pi)^2 + \pi(1 - \pi)\frac{8}{9}\right](1 - \lambda)$$

from being truthful whereas

$$\mathbb{E}[U_3^S(1-\theta_3;\theta_3,\lambda)|h_{30}] = -\lambda - \left[\frac{2\pi(1-\pi)\varepsilon_3}{9}\right](1-\lambda)$$

from falsifying. Therefore, she falsifies if $\lambda < \lambda_3(\varepsilon_3, h_{30})$, where

$$\lambda_3(\varepsilon_3, h_{30}) := \frac{9 - 2\pi(1-\pi)(5+\varepsilon_3)}{18 - 2\pi(1-\pi)(5+\varepsilon_3)}.$$

The less complex the heuristic, the more steeply the cut-off is increasing in π. The impact of the complexity of the heuristic is most pronounced in uninformative environments as $\pi \to 1/2$. The impact becomes negligible as $\pi \to 1$.

Given that we assume player 3 can infer player 2's type when $\lambda_2 = \lambda$, we now find the lower bound on λ such that $\mathbb{E}[U_2^S(\theta_2;\theta_2,\lambda)|h_{20}] \geqslant \mathbb{E}[U_2^S(1-\theta_2;\theta_2,\lambda)|h_{20}]$ at $h_{20} = \{1-\theta_2\}$. Moreover, we presume (and verify below) this lower bound to be weakly above 1/4. Hence, when player 2 is truthful so that the history will be "mixed" for player 4 (i.e., $\{1-\theta_2, \theta_2, 1-\theta_2\}$ or $\{1-\theta_2, \theta_2, \theta_2\}$), player 4 is predicted to always be truthful. Then, the analysis of player 2 is as in **Case I(a)** with player 2 being truthful if $\lambda \geqslant \lambda_2(\varepsilon_2, h_{20}) \geqslant 1/4$, which verifies our presumption. Using the above results together with $\lambda_2(\varepsilon_2, h_{20}) < \lambda_3(\varepsilon_3, h_{30}) < 1/2$, the mapping is $(\theta,\theta,\theta_3,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ and $(\theta, 1-\theta, \theta_3, \theta_4) \mapsto (\theta, 1-\theta, \theta_3, \theta_4)$ if $\lambda \in \big[\lambda_2(\varepsilon_2, h_{20}), \lambda_3(\varepsilon_3, h_{30})\big)$ whereas $(\theta,\theta,\theta,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ and $\{(\theta,\theta,1-\theta,\theta_4), (\theta,1-\theta,\theta,\theta_4), (\theta,1-\theta,1-\theta,\theta_4)\} \mapsto (\theta,\theta_2,\theta_3,\theta_4)$ if $\lambda \in \big[\lambda_3(\varepsilon_3, h_{30}), 1/2\big)$. □

A.3.3 Heterogeneity in the Complexity of Heuristics

Given Definition 1 and Definition 2, allowing for heterogeneity in the complexity of heuristics does not change a credulous or sceptical player's cut-off, because either the complexity of player j's heuristic is irrelevant information to player $i \neq j$ (whenever $\hat{\lambda}_{ij} = 1$), or the presumption about the complexity of player j's heuristic matters, not the actual complexity. Specifically, a sceptical player continues to presume her predecessors to be credulous (Footnote 9).

Moreover, heterogeneity in the complexity of heuristics does not change our result that a change from credulity to rationality increases the ex-ante probability of uniform behaviour whilst a change to scepticism decreases it. Therefore, this result is robust to switching one interior player or the whole group. In particular, in the group of four players with a homogeneous conformity motive but heterogeneous heuristics, Theorem 1 generalises to:

Theorem 1″. *The mapping between types and actions is:*

(i) $(\theta,\theta_2,\theta_3,\theta_4) \mapsto (\theta,\theta,\theta,\theta)$ *if* $\lambda \in \big[0, \underline{\lambda}\big)$,

(ii) $(\theta, \theta, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $(\theta, 1-\theta, \theta_3, \theta_4) \mapsto (\theta, 1-\theta, \theta_3, \theta_4)$ *if* $\lambda \in [\underline{\lambda}, \bar{\lambda})$,

(iii) $(\theta, \theta, \theta, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ *and* $\{(\theta, \theta, 1-\theta, \theta_4), (\theta, 1-\theta, \theta, \theta_4), (\theta, 1-\theta, 1-\theta, \theta_4)\} \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if* $\lambda \in [\bar{\lambda}, 1/2)$,

(iv) $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta_2, \theta_3, \theta_4)$ *if* $\lambda \in [1/2, 1]$,

where $\underline{\lambda} \in \{\lambda_2(0, \theta_2), \lambda_2^R(h_{20}, \theta_2)\}$ *and* $\bar{\lambda} \in \{\lambda_3(0, \theta_3), \lambda_3(h_{30}^S, \theta_3), \lambda_3^R(h_{30}, \theta_3)\}$.

That is, the complexity of player 2's heuristic affects the upper bound of mapping (i) and the lower bound of mapping (ii) whilst player 3's heuristic affects the upper bound of mapping (ii) and the lower bound of mapping (iii) in Theorem 1″. Hence, whenever the true conformity motive λ has a distribution F with full support, a targeted rationality boost of an interior player (i.e., player 2 or 3) shifts probability mass to a mapping in which uniform behaviour is more likely. A rationality boost of the whole group is, thus, not necessary for Corollary 1 to hold.

Moreover, Theorem 1″ highlights that a policy maker who faces a group of credulous players does not need to boost the whole group to achieve the maximum effect: a policy maker concerned with reducing (increasing) the ex-ante probability of uniform behaviour achieves the maximum change by boosting player 3 (players 2 and 3). The effectiveness of a targeted, relative to an aggregate, boost may be important when its implementation is costly. In particular, "boosts often require investments in time, effort, and motivation" (Hertwig and Grüne-Yanoff, 2017, p.982). To ensure take-up notwithstanding, boosts have to be explicit, visible, and transparent, which often prevents them from being no-cost interventions (ibid.).

Proof of Theorem 1″. Let $N = 4$ and the common conformity motive be λ. Given that (i) player 1 is always truthful, (ii) the sceptical and credulous player 2 have identical cut-offs, and (iii) player 4's cut-off is independent of the complexity of her heuristic, there are six different combinations of credulity, scepticism and rationality (i.e. 2×3) to consider:

[1] $(\cdot, C, C, \cdot) = (\cdot, S, C, \cdot)$,

[2] $(\cdot, S, S, \cdot) = (\cdot, C, S, \cdot)$,

[3] $(\cdot, R, R, \cdot)$,

[4] $(\cdot, C, R, \cdot) = (\cdot, S, R, \cdot)$,

[5] $(\cdot, R, C, \cdot)$,

[6] $(\cdot, R, S, \cdot)$.

We analysed **Cases 1, 2 and 3** in the proof of Theorem 1. We now analyse **Cases 4, 5 and 6**.

Case 4 The rational player 3 understands when she can infer θ_2 from player 2's action. This does not depend on whether player 2 is credulous, sceptical or rational. Hence, player 3's cut-off is as in Table B.2. The credulous player 2 correctly infers that player 1 is truthful but does not realise that she potentially has a strategic externality on her successors. In particular, as $\hat{\lambda}_{23} = \hat{\lambda}_{24} = 1$, player 2's cut-off is $\lambda_2(0, \theta_2) = 1/4$. Following the same reasoning as in the proof of Theorem 1, mapping (i) obtains for $\lambda \in \big[0, \lambda_2(0, \theta_2)\big)$, mapping (ii) obtains for $\lambda \in \big[\lambda_2(0, \theta_2), \lambda_3^R(h_{30}, \theta_3)\big)$, mapping (iii) obtains for $\lambda \in \big[\lambda_3^R(h_{30}, \theta_3), 1/2\big)$, and mapping (iv) obtains for $\lambda \in [1/2, 1]$, where $\lambda_3^R(h_{30}, \theta_3) = (9-10\pi(1-\pi))/(18-10\pi(1-\pi))$.

Case 5 Player 3 is credulous and, thus, has cut-off $\lambda_3(0, \theta_3) = (3-4\pi(1-\pi))/(6-4\pi(1-\pi))$ irrespective of whether player 2 was, in fact, truthful. Player 2 is rational and, hence, takes into account that player 3 posits $\hat{\lambda}_{32} = 1$. So player 2 takes into account that player 3 presumes s_2 reveals θ_2. In addition to correctly inferring player 3's cut-off $\lambda_3(0, \theta_3)$, the rational player 2 takes into account player 3's true conformity motive: $\hat{\lambda}_{23} = \lambda$. This means that, as in the all-rational group, player 2 correctly predicts player 3's behaviour. In particular, if (i) $\lambda \in [0, 1/4)$, player 3 and 4 agree with player 2 when $s_2 = s_1$; player 3 is truthful and player 4 agrees with whatever player 3 announces otherwise. If (ii) $\lambda \in [1/4, \lambda_3(0, \theta_3))$, player 3 and 4 agree with player 2 when $s_2 = s_1$; player 3 and 4 are truthful when $s_2 \neq s_1$. If (iii) $\lambda \in [\lambda_3(0, \theta_3), 1/2)$, player 3 is always truthful but player 4 agrees with everyone else when $s_1 = s_2 = s_3$. If (iv) $\lambda \in [1/2, 1]$, player 3 and 4 are always truthful.

Finally, we calculate player 2's cut-off for each of the four cases above to check that in (i) player 2's cut-off is greater and in (ii) player 2's cut-off is lower than the upper bound. This implies that for $\lambda \in \big[0, \lambda_2^R(h_{20}, \theta_2)\big)$, mapping (i) of Theorem 1 obtains, whilst for $\lambda \in \big[\lambda_2^R(h_{20}, \theta_2), \lambda_3(0, \theta_3)\big)$, mapping (ii) of Theorem 1 obtains, where $\lambda_2^R(h_{20}, \theta_2) = (5 - 2\pi(1-\pi))/(14 - 2\pi(1-\pi))$. Because player 2's cut-off in case (iii) is lower than the lower bound of the interval, for $\lambda \in [\lambda_3(0, \theta_3), 1/2)$, mapping (iii) of Theorem 1 obtains.

Case 6 As explained in Footnote 9, the sceptical player 3 implicitly presumes player 2 to be credulous irrespective of the complexity of her heuristic. Hence, when player 3 observes $h_{30} = \{1-\theta_3, 1-\theta_3\}$, she edits the history to $h_{30}^S = \{1-\theta_3\}$ with corresponding cut-off $\lambda_3(h_{30}^S, \theta_3) = 2/5$. When instead the history is mixed, $\{1-\theta_3, \theta_3\}$ or $\{\theta_3, 1-\theta_3\}$, she does not edit and is truthful. Similarly, at $h_{31} = \{\theta_3, \theta_3\}$ the group size is too small to possibly induce falsification after editing.

When the rational player 2 takes into account that she faces a sceptical player who posits $\hat{\lambda}_{32} = 0$ (or sufficiently low) whenever $s_2 = s_1$, she takes into account that player

3's cut-off at h_{30} is lower compared to **Case 5**: $2/5 < (3 - 4\pi(1-\pi))/(6 - 4\pi(1-\pi))$. Apart from this difference, the analysis for player 2 is as in **Case 5** because $\lambda_2^R(h_{20}, \theta_2) < \lambda_3(h_{30}^S, \theta_3)$. This implies that for $\lambda \in \big[0, \lambda_2^R(h_{20}, \theta_2)\big)$, mapping (i) of Theorem 1 obtains, whilst for $\lambda \in \big[\lambda_2^R(h_{20}, \theta_2), \lambda_3(h_{30}^S, \theta_3)\big)$, mapping (ii) obtains. Because player 2's cut-off in case (iii) is lower than the lower bound of the interval, for $\lambda \in \big[\lambda_3(h_{30}^S, \theta_3), 1/2\big)$, mapping (iii) obtains. □

A.3.4 Additional Results

Additional Result 1 (Informativeness of type)**.** *In period $i \in \{1, 2, \ldots, N-1\}$, if*

C *(i) $M_i > (i-2)/2$, then $\partial\lambda_i(M_i, \theta_i)/\partial\pi < 0$; or,*

(ii) $M_i = (i-2)/2$, then $\partial\lambda_i(M_i, \theta_i)/\partial\pi = 0$; or,

(iii) $M_i < (i-2)/2$, then $\partial\lambda_i(M_i, \theta_i)/\partial\pi > 0$.

S *(i) $M_i > (i - 2 + A_i^{=} - A_i^{\neq})/2$, then $\partial\lambda_i(h_i^S, \theta_i)/\partial\pi < 0$; or,*

(ii) $M_i = (i - 2 + A_i^{=} - A_i^{\neq})/2$, then $\partial\lambda_i(h_i^S, \theta_i)/\partial\pi = 0$; or,

(iii) $M_i < (i - 2 + A_i^{=} - A_i^{\neq})/2$, then $\partial\lambda_i(h_i^S, \theta_i)/\partial\pi > 0$.

Proof of Additional Result 1. To prove Additional Result 1, we use the fact that $\partial\lambda_i(\cdot, 0)/\partial\pi \overset{\text{sgn}}{=} \partial\mathbb{E}[\bar{s}_{-i}|\cdot, 0]/\partial\pi$.

C *(i)* Suppose $\theta_i = 0$ and that, including player i, there is a strict majority for θ_i, which implies $(M_i+1)/i > 1/2$. Then, by (B.12), the sign of $\partial\mathbb{E}[\bar{s}_{-i}|M_i, 0]/\partial\pi$ is determined by

$$\frac{\partial}{\partial\pi}\left[\frac{(1-\pi)^a\pi + \pi^a(1-\pi)}{\pi^a + (1-\pi)^a}\right] = \frac{\pi(1-\pi)[\pi^{2a} - (1-\pi)^{2a}] + a(2\pi-1)[\pi(1-\pi)]^a}{\pi(\pi-1)[\pi^a + (1-\pi)^a]^2}, \quad \text{(A.26)}$$

where $a := 2(M_i+1) - i$. The assumption of a strict majority implies that $a > 0$. Since $(\pi - 1) < 0$, (A.26) is negative if its numerator is positive. Since $(2\pi - 1) > 0$, the assumption that $a > 0$ is sufficient for the numerator to be strictly positive as $\pi^{2a} - (1-\pi)^{2a} > 0$ for all $\pi \in (1/2, 1)$ if $a > 0$. Finally, note that $a > 0$ implies that $2(M_i+1) - i > 0$, which rearranges to $M_i > (i-2)/2$.

C *(ii)* Suppose $\theta_i = 0$ and that, including player i, there is a tie, which implies $\Pr(\omega = \theta_i|M_i, \theta_i) = 1/2$. Then,

$$\mathbb{E}[\bar{s}_i|M_i, 0] = \frac{i-1-M_i}{N-1} + \left[\frac{\pi}{2} + \frac{(1-\pi)}{2}\right]\left(\frac{N-i}{N-1}\right) = \frac{i-1-M_i}{N-1} + \frac{N-i}{2(N-1)},$$

which is independent of π implying that $\partial\lambda_i(M_i, 0)/\partial\pi = 0$. Finally, note that for a tie, we require that $M_i = (i-2)/2$.

C (iii) Suppose $\theta_i = 0$ and that, including player i, there is a strict minority for θ_i, which implies $(M_i + 1)/i < 1/2$. Then, by (B.13), the sign of $\partial\mathbb{E}[\bar{s}_{-i}|M_i, 0]/\partial\pi$ is determined by

$$\frac{\partial}{\partial\pi}\left[\frac{(1-\pi)^{1-a}+\pi^{1-a}}{(1-\pi)^{-a}+\pi^{-a}}\right] = \frac{\pi(1-\pi)[\pi^{2a}-(1-\pi)^{2a}]+a(2\pi-1)[\pi(1-\pi)]^{a}}{\pi(\pi-1)[\pi^{a}+(1-\pi)^{a}]^{2}}. \tag{A.27}$$

The assumption of a strict minority implies that $a < 0$. Since $(\pi - 1) < 0$, (A.27) is positive if its numerator is negative. Since $(2\pi - 1) > 0$, the assumption that $a < 0$ is sufficient for the numerator to be strictly negative as $\pi^{2a} - (1 - \pi)^{2a} < 0$ for all $\pi \in (1/2, 1)$ if $a < 0$. Finally, note that $a < 0$ implies that $2(M_i + 1) - i < 0$, which rearranges to $M_i < (i - 2)/2$.

S (i) Suppose $\theta_i = 0$ and that, including i, there is a strict majority for θ_i, which implies $M_i > (i - 2 + A_i^{=} - A_i^{\neq})/2$. Then,

$$\mathbb{E}[\bar{s}_{-i}|h_i^S, 0] = \frac{i-1-M_i}{N-1} + \frac{(1-\pi)^c\pi + \pi^c(1-\pi)}{\pi^c + (1-\pi)^c}\left(\frac{N-i}{N-1}\right),$$

where $c := 2(M_i + 1) - i + A_i^{\neq} - A_i^{=}$. Therefore, the sign of $\partial\mathbb{E}[\bar{s}_{-i}|h_i^S, 0]/\partial\pi$ is determined by (A.26) with a replaced by c. The assumption of a strict majority implies $c > 0$, or equivalently $M_i > (i - 2 + A_i^{=} - A_i^{\neq})/2$.

S (ii) The proof follows the same logic as in C (ii).

S (iii) Suppose $\theta_i = 0$ and that, including i, there is a strict minority for θ_i, which implies $M_i < (i - 2 + A_i^{=} - A_i^{\neq})/2$. Then,

$$\mathbb{E}[\bar{s}_{-i}|h_i^S, 0] = \frac{i-1-M_i}{N-1} + \frac{(1-\pi)^{1-c}+\pi^{1-c}}{(1-\pi)^{-c}+\pi^{-c}}\left(\frac{N-i}{N-1}\right).$$

Therefore, the sign of $\partial\mathbb{E}[\bar{s}_{-i}|h_i^S, 0]/\partial\pi$ is determined by (A.27) with a replaced by c. The assumption of a strict minority implies $c < 0$, or equivalently $M_i < (i - 2 + A_i^{=} - A_i^{\neq})/2$. □

Appendix B

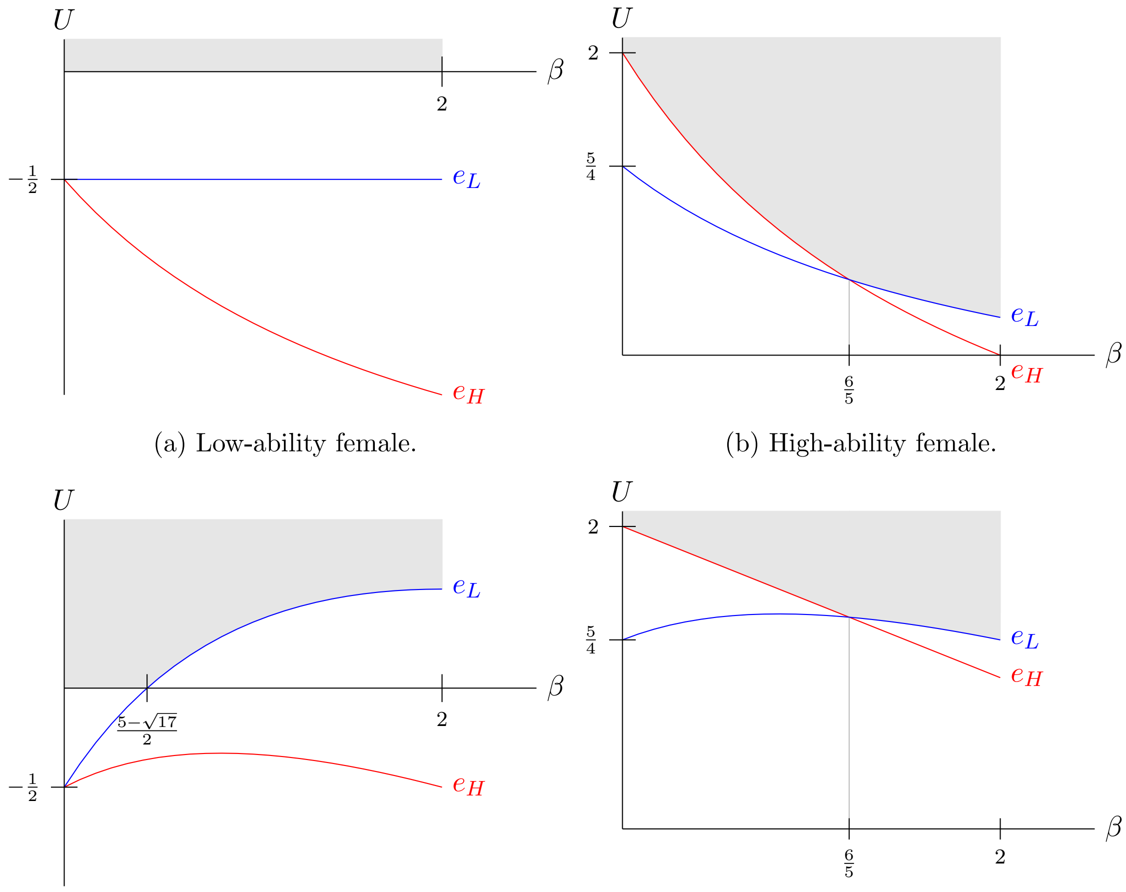
Proofs

B.1 Chapter 1 Proofs

Proof of Lemma 1. First, note that substituting for the hiring probabilities in equation (2.2) yield

$$
\begin{aligned}
U(e_L, \eta_L, f) &= -\frac{1}{2}, & U(e_L, \eta_L, m) &= \frac{\beta}{2+\beta}w - \frac{1}{2}, \\
U(e_H, \eta_L, f) &= \frac{1}{2+\beta}w - 2, & U(e_H, \eta_L, m) &= \frac{1+\beta}{2+\beta}w - 2, \\
U(e_L, \eta_H, f) &= \frac{1}{2+\beta}w - \frac{1}{4}, & U(e_L, \eta_H, m) &= \frac{1+\beta}{2+\beta}w - \frac{1}{4}, \\
U(e_H, \eta_H, f) &= \frac{2}{2+\beta}w - 1, & U(e_H, \eta_H, m) &= w - 1
\end{aligned}
$$

in an informed audition, where by Assumption 3 an applicant's wage is $w = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$.



(a) Low-ability female.

(b) High-ability female.

(c) Low-ability male.

(d) High-ability male.

Figure B.1: Effort choices in an informed audition.

Low-ability females drop out of the audition for all possible biases (Figure B.1a). High-ability females exert high effort if $\beta \in [0, \frac{6}{5}]$ and low effort if $\beta \in (\frac{6}{5}, 2]$ (Figure B.1b). Low-ability males drop out if β is strictly lower than $\frac{5-\sqrt{17}}{2}$, and exert low effort if $\beta \in [\frac{5-\sqrt{17}}{2}, 2]$ (Figure B.1c). High-ability males exert high effort if $\beta \in [0, \frac{6}{5}]$ and low effort if $\beta \in (\frac{6}{5}, 2]$ (Figure B.1d). Note the gender differences for low-ability applicants: $U(e_L, \eta_L, m)$ is increasing in the evaluator's bias while $U(e_L, \eta_L, f)$ is negative and independent of the bias. □

Proof of Proposition 1. To derive the evaluator's expected net utility when his bias is low, moderate, or high, I first determine his gross utility in each of the four possible subgames. Using the fact that subgames (η_L, f) and (η_L, m) are each reached with probability $(1 - p)/2$ whilst (η_H, f) and (η_H, m) are each reached with probability $p/2$, I then derive the evaluator's overall expected gross utility in stage 1. Only in the last step, I subtract wage costs as $w := w_\theta = w_{\bar{U}}$ for all $\theta \in \Theta$ by Assumption 3.

Suppose the evaluator's bias against female applicants is low; that is, $\beta \in \beta_I^L$ where β_I^L is defined as in Lemma 1(i). Then, low-ability males and females do not participate and high-ability males and females exert high effort. In subgame (η_L, f) and (η_L, m), the evaluator, therefore, hires the outside option with probability one. With a slight abuse of notation, the evaluator's expected gross utility is the expected value of the outside option:

$$\mathbb{E}[V|O, \eta_L, f] = \mathbb{E}[V|O, \eta_L, m] = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}.$$

As $w = \mathbb{E}[\bar{U}]$, the evaluator, by construction, breaks even after accounting for wage costs in these two subgames. In subgame (η_H, f), the evaluator's gross utility from hiring the applicant is $4-\beta$ with corresponding hiring probability $\frac{2}{2+\beta}$. The evaluator hires the outside option with expected value $3 - \frac{\beta}{2}$ with complementary probability $\frac{\beta}{2+\beta}$. The evaluator's expected gross utility in subgame (η_H, f) can then be calculated as the probability that he hires the applicant times the utility from hiring her plus the probability that he hires the outside option times the expected value of the outside option:[1]

$$\begin{aligned}\mathbb{E}[V|q_H, \eta_H, f] &= \Pr(h|q_H, \eta_H, f)V(q_H, \eta_H, f) + [1 - \Pr(h|q_H, \eta_H, f)]\mathbb{E}[\bar{U}], \\ &= \frac{2}{2+\beta}(4-\beta) + \frac{\beta}{2+\beta}\left(3 - \frac{\beta}{2}\right) = 2 - \frac{\beta}{2} + \frac{4}{2+\beta}.\end{aligned}$$

In subgame (η_H, m), the evaluator's gross utility from hiring the applicant is 4 with corresponding hiring probability 1. His expected gross utility in this subgame, therefore, is

$$\mathbb{E}[V|q_H, \eta_H, m] = \Pr(h|q_H, \eta_H, m)V(q_H, \eta_H, m) = 4.$$

Under the common prior assumption, the evaluator's overall expected gross utility in stage 1 in an informed audition is

$$\begin{aligned}\mathbb{E}[V|\beta] &= \frac{1-p}{2}\Big[\mathbb{E}[V|O, \eta_L, f] + \mathbb{E}[V|O, \eta_L, m]\Big] + \frac{p}{2}\Big[\mathbb{E}[V|q_H, \eta_H, f] + \mathbb{E}[V|q_H, \eta_H, m]\Big], \\ &= \frac{1-p}{2}\left[\left(3 - \frac{\beta}{2}\right) + \left(3 - \frac{\beta}{2}\right)\right] + \frac{p}{2}\left[\left(2 - \frac{\beta}{2} + \frac{4}{2+\beta}\right) + 4\right].\end{aligned}$$

After subtracting wage costs, the evaluator's expected net utility is

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^L] := \mathbb{E}[V|\beta \in \beta_I^L] - w = \frac{p}{2}\left[\frac{2}{2+\beta}\left(1 - \frac{\beta}{2}\right) + \left(1 + \frac{\beta}{2}\right)\right] > 0.$$

Suppose the evaluator is moderately biased against female applicants; that is, $\beta \in \beta_I^M$ where β_I^M is defined as in Lemma 1(ii). Then, subgame (η_L, f), (η_H, f) and (η_H, m) remain

[1]Recall that, by Assumption 2, the value of $\bar{U}$ is unknown to the evaluator when making the hiring decision. Therefore, all calculations use the *unconditional* expected value of the outside option. The model's results are robust to altering the timing and using the *conditional* expected value. However, when engaging a substitute from an agency, it is more appropriately interpreted as a gamble as done here.

unchanged. Low-ability males participate and exert low effort. The evaluator's gross utility from hiring the applicant in subgame (c) changes to 2 with corresponding hiring probability $\frac{\beta}{2+\beta}$. The evaluator hires the outside option with expected value $3-\frac{\beta}{2}$ with complementary probability $\frac{2}{2+\beta}$. The evaluator's expected gross utility in subgame (η_L, m), therefore, becomes

$$\begin{aligned}\mathbb{E}[V|q_L,\eta_L,m] &= \Pr(h|q_L,\eta_L,m)V(q_L,\eta_L,m) + [1-\Pr(h|q_L,\eta_L,m)]\mathbb{E}[\bar{U}],\\ &= \frac{\beta}{2+\beta}2 + \frac{2}{2+\beta}\left(3-\frac{\beta}{2}\right) = 1 + \frac{4}{2+\beta},\end{aligned}$$

which is strictly less than $\mathbb{E}[\bar{U}]$ for all β. Because $w = \mathbb{E}[\bar{U}]$, the evaluator's expected utility net of wage costs is, thus, negative in subgame (η_L, m) when he is moderately biased. The evaluator's overall expected gross utility becomes $\mathbb{E}[V|\beta \in \beta_I^M]$

$$\begin{aligned}&= \frac{1-p}{2}\Big[\mathbb{E}[V|O,\eta_L,f] + \mathbb{E}[V|q_L,\eta_L,m]\Big] + \frac{p}{2}\Big[\mathbb{E}[V|q_H,\eta_H,f] + \mathbb{E}[V|q_H,\eta_H,m]\Big],\\ &= \frac{1-p}{2}\left[\left(3-\frac{\beta}{2}\right) + \left(1+\frac{4}{2+\beta}\right)\right] + \frac{p}{2}\left[\left(2-\frac{\beta}{2}+\frac{2}{2+\beta}\right)+4\right].\end{aligned}$$

After subtracting wage costs, the evaluator's expected net utility is

$$\begin{aligned}\mathbb{E}[\Pi_I|\beta \in \beta_I^M] &:= \mathbb{E}[V|\beta \in \beta_I^M] - w,\\ &= \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right) + \left(1+\frac{\beta}{2}\right)\right] \gtrless 0.\end{aligned}$$

Suppose the evaluator is highly biased against female applicants; that is, $\beta \in \beta_I^H$ where β_I^H is defined as in Lemma 1(iii). Then, subgame (η_L, f) and (η_L, m) remain unchanged. High-ability females and males exert low effort. The evaluator's gross utility from hiring the applicant in subgame (η_H, f) changes to $3-\beta$ with corresponding hiring probability $\frac{1}{2+\beta}$. The evaluator hires the outside option with expected value $3-\frac{\beta}{2}$ with complementary probability $\frac{1+\beta}{2+\beta}$. The evaluator's expected gross utility in subgame (η_H, f), therefore, falls to

$$\begin{aligned}\mathbb{E}[V|q_L,\eta_H,f] &= \Pr(h|q_L,\eta_H,f)V(q_L,\eta_H,f) + [1-\Pr(h|q_L,\eta_H,f)]\mathbb{E}[\bar{U}],\\ &= \frac{1}{2+\beta}(3-\beta) + \frac{1+\beta}{2+\beta}\left(3-\frac{\beta}{2}\right) = \frac{5}{2} - \frac{\beta}{2} + \frac{1}{2+\beta}.\end{aligned}$$

The evaluator's gross utility from hiring the applicant in subgame (η_H, m) changes to 3 with corresponding hiring probability $\frac{1+\beta}{2+\beta}$. The evaluator hires the outside option with expected value $3-\frac{\beta}{2}$ with complementary probability $\frac{1}{2+\beta}$. The evaluator's expected gross utility in subgame (η_H, m), therefore, falls to

$$\begin{aligned}\mathbb{E}[V|q_L,\eta_H,m] &= \Pr(h|q_L,\eta_H,m)V(q_L,\eta_H,m) + [1-\Pr(h|q_L,\eta_H,m)]\mathbb{E}[\bar{U}],\\ &= \frac{1+\beta}{2+\beta}3 + \frac{1}{2+\beta}\left(3-\frac{\beta}{2}\right) = \frac{5}{2} + \frac{1}{2+\beta}.\end{aligned}$$

The evaluator's overall expected gross utility becomes $\mathbb{E}[V|\beta \in \beta_I^H]$

$$= \frac{1-p}{2}\Big[\mathbb{E}[V|O,\eta_L,f] + \mathbb{E}[V|q_L,\eta_L,m]\Big] + \frac{p}{2}\Big[\mathbb{E}[V|q_L,\eta_H,f] + \mathbb{E}[V|q_L,\eta_H,m]\Big],$$
$$= \frac{1-p}{2}\left[\left(3-\frac{\beta}{2}\right) + \left(1+\frac{4}{2+\beta}\right)\right] + \frac{p}{2}\left[\left(\frac{5}{2}-\frac{\beta}{2}+\frac{1}{2+\beta}\right) + \left(\frac{5}{2}+\frac{1}{2+\beta}\right)\right].$$

After subtracting wage costs, the evaluator's expected net utility is

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^H] := \mathbb{E}[V|\beta \in \beta_I^H] - w,$$
$$= \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{1}{2+\beta}\left(-\frac{\beta}{2}\right) + \frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)\right] \gtreqless 0.$$

□

Proof of Lemma 2. First, note that substituting for the two hiring probabilities in equation (2.2) yield

$$U(e_L,\eta_L) = \frac{\beta}{4+2\beta}w - \frac{1}{2}, \qquad U(e_H,\eta_L) = \frac{2(1+p)+\beta}{4+2\beta}w - 2,$$
$$U(e_L,\eta_H) = \frac{\beta}{4+2\beta}w - \frac{1}{4}, \qquad U(e_H,\eta_H) = \frac{2(1+p)+\beta}{4+2\beta}w - 1$$

for both female and male applicants in a pooling blind audition, where $w = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}$.

(i) For $p < \frac{1}{3}$, I show that $U(e_H,\eta_L)$ is strictly negative for all possible values of β. It is always better for low-ability applicants to not participate. First, note that $U(e_H,\eta_L)$ is strictly decreasing in β:

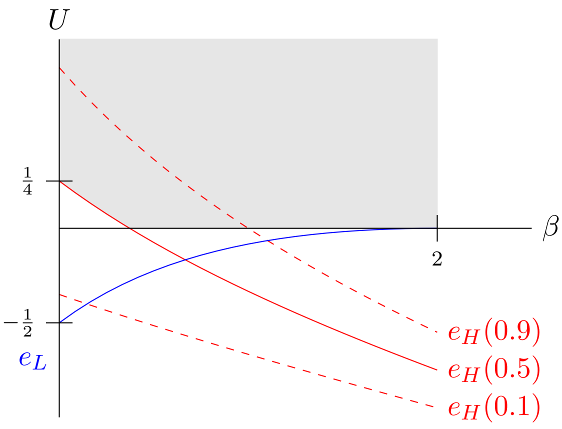
$$\frac{\partial U(e_H,\eta_L)}{\partial \beta} = -\frac{4p}{(4+2\beta)^2}\left(3-\frac{\beta}{2}\right) - \frac{2(1+p)+\beta}{8+4\beta} < 0 \ \forall\ \beta \in [0,2].$$

This implies that, if $U(e_H,\eta_L) < 0$ at $\beta = 0$, then $U(e_H,\eta_L) < 0$ for all $\beta \in [0,2]$. I can use this condition to solve for the corresponding prior:

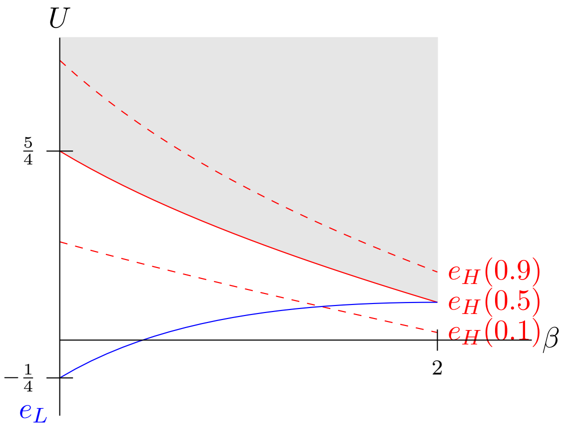
$$U(e_H,\eta_L)\Big|_{\beta=0} = \frac{1+p}{2}3 - 2 < 0 \Rightarrow p < \frac{1}{3}.$$

Thus, for $p < \frac{1}{3}$, pooling on high effort is never possible.

(ii) For $p = \frac{1}{3}$, $U(e_H,\eta_L) = 0$ at $\beta = 0$. By the previous argument, $U(e_H,\eta_L) < 0$ for all $\beta \in (0,2]$. This implies that low-ability applicants are indifferent between exerting high effort and not participating only if the evaluator is unbiased. Furthermore, it is optimal for high-ability applicants to exert high effort whenever it is optimal for low-ability applicants to exert high effort. This is because $U(e_L,\cdot)$ shifts up by $\frac{1}{4}$ while $U(e_H,\cdot)$ shifts up by 1 in the move from η_L to η_H (Figure B.2a and Figure B.2b, respectively). Thus, pooling on high effort is possible at $p = \frac{1}{3}$ if $\beta = 0$.



(a) Low-ability applicant.



(b) High-ability applicant.

Figure B.2: Effort choices in a pooling blind audition for different priors.

(iii) For $\frac{1}{3} < p < 1$, there exists a range of biases for which $U(e_H, \eta_L)$ is positive. The upper bound on β for which a pooling equilibrium can be supported is determined by the indifference condition

$$U(e_H, \eta_L) = \frac{2(1+p)+\beta}{4+2\beta}\left(3 - \frac{\beta}{2}\right) - 2 = 0.$$

Given that $\beta \in [0, 2]$, the above equation is uniquely solved by $\beta(p) := \sqrt{p^2 + 16p} - p - 2$. The upper bound $\beta(p)$ is strictly increasing in p:

$$\frac{d\beta(p)}{dp} = \frac{p+8}{\sqrt{p(p+16)}} - 1 > 0 \ \forall \ 0 < p < 1.$$

Taking the limit of $\beta(p)$ yields

$$\lim_{p \uparrow 1} \beta(p) = \sqrt{17} - 3 > 0.$$

□

Proof of Proposition 2. Suppose the evaluator's bias against female applicants is low; that is, $\beta \in \beta_B^L$ with $\frac{1}{3} \leqslant p < 1$ where β_B^L is defined as in Lemma 2. His expected utility from hiring the applicant at the information set q_H is $3 + p - \frac{\beta}{2}$ with corresponding hiring probability $\frac{2(1+p)+\beta}{4+2\beta}$. The evaluator hires the outside option with expected value $3 - \frac{\beta}{2}$ with complementary probability $\frac{2(1-p)+\beta}{4+2\beta}$. Given the evaluator's beliefs, his overall expected gross utility in stage 1 in a pooling blind audition is

$$\begin{aligned}\mathbb{E}[V|\beta \in \beta_B^L] &= \Pr(h|q_H)\mathbb{E}[V|q_H] + [1 - \Pr(h|q_H)]\mathbb{E}[\bar{U}],\\ &= \frac{2(1+p)+\beta}{4+2\beta}\left(3 + p - \frac{\beta}{2}\right) + \frac{2(1-p)+\beta}{4+2\beta}\left(3 - \frac{\beta}{2}\right).\end{aligned}$$

After subtracting wage costs, the evaluator's expected net utility is

$$\mathbb{E}[\Pi_B|\beta \in \beta_B^L] := \mathbb{E}[V|\beta \in \beta_B^L] - w,$$
$$= \frac{2(1+p)+\beta}{4+2\beta}\left[3+p-\frac{\beta}{2}-\left(3-\frac{\beta}{2}\right)\right] = p\frac{2(1+p)+\beta}{4+2\beta} > 0.$$

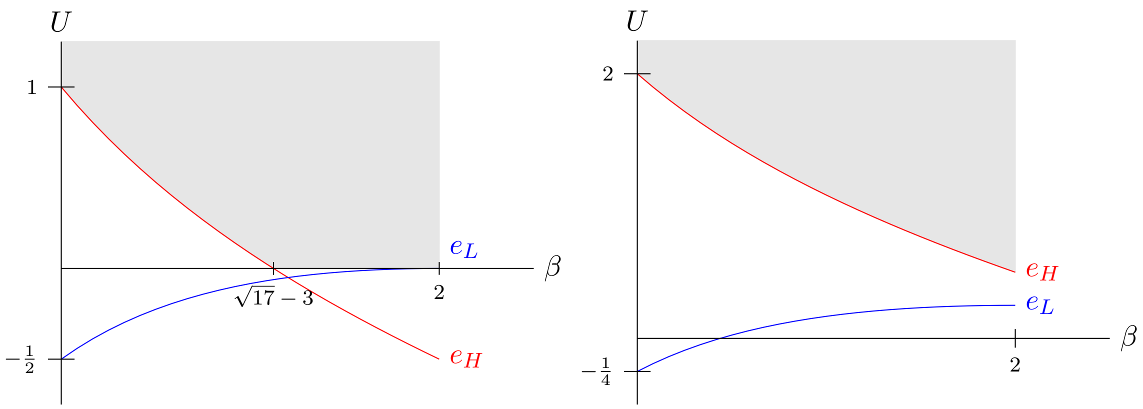
□

Proof of Lemma 3. First, note that substituting for the two hiring probabilities in the applicant's expected utility (2.2) yield

$$U(e_L,\eta_L) = \frac{\beta}{4+2\beta}w - \frac{1}{2}, \qquad U(e_H,\eta_L) = \frac{4+\beta}{4+2\beta}w - 2,$$
$$U(e_L,\eta_H) = \frac{\beta}{4+2\beta}w - \frac{1}{4}, \qquad U(e_H,\eta_H) = \frac{4+\beta}{4+2\beta}w - 1$$

for both male and female applicants in a separating blind audition, where $w = \mathbb{E}[\bar{U}] = 3-\frac{\beta}{2}$.

For $\beta \in [0, \sqrt{17}-3]$, low-ability applicants exert high effort (Figure B.3a). For $\beta \in (\sqrt{17}-3, 2)$, low-ability applicants do not participate in the audition (Figure B.3a). For $\beta \in [0,2]$, high-ability applicants exert high effort in the audition (Figure B.3b).



(a) Low-ability applicant. (b) High-ability applicant.

Figure B.3: Effort choices in a separating blind audition.

□

Proof of Proposition 3. Suppose the evaluator's bias against female applicants is high; that is $\beta \in \beta_B^H$ where β_B^H is defined as in Lemma 3. The evaluator's expected utility from hiring the applicant at the information set q_H is $4-\frac{\beta}{2}$ with corresponding hiring probability $\frac{4+\beta}{4+2\beta}$. The evaluator hires the outside option with expected value $3-\frac{\beta}{2}$ with complementary probability $\frac{\beta}{2+\beta}$. If the applicant is a low-ability male or female, the evaluator hires the outside option with probability one as those types do not participate. In this case, the evaluator's expected utility is the expected value of the outside option:

$$\mathbb{E}[V|O] = \mathbb{E}[\bar{U}] = 3 - \frac{\beta}{2}.$$

As $w = \mathbb{E}[\bar{U}]$, the evaluator breaks even by construction after accounting for wage costs. Given the evaluator's beliefs, his overall expected gross utility in stage 1 in a separating blind audition is

$$\begin{aligned}\mathbb{E}[V|\beta \in \beta_B^H] &= p\Bigg[\Pr(h|q_H)\mathbb{E}[V|q_H] + [1 - \Pr(h|q_H)]\mathbb{E}[\bar{U}]\Bigg] + (1-p)\Bigg[\mathbb{E}[\bar{U}]\Bigg],\\ &= p\left[\frac{4+\beta}{4+2\beta}\left(4 - \frac{\beta}{2}\right) + \frac{\beta}{2+\beta}\left(3 - \frac{\beta}{2}\right)\right] + (1-p)\left[3 - \frac{\beta}{2}\right].\end{aligned}$$

After subtracting wage costs, the evaluator's expected net utility is

$$\begin{aligned}\mathbb{E}[\Pi_B|\beta\beta_B^H] &:= \mathbb{E}[V|\beta \in \beta_B^H] - w,\\ &= p\left[\frac{4+\beta}{4+2\beta}\left(4 - \frac{\beta}{2} - \left(3 - \frac{\beta}{2}\right)\right)\right] = p\frac{4+\beta}{4+2\beta} > 0.\end{aligned}$$

□

Proof of Proposition 4. (i) For $\beta \in \beta_B^L := [0, \sqrt{p^2 + 16p} - p - 2]$, I show that the evaluator's expected net utility in a blind audition is strictly lower than his expected net utility in an informed audition. First, from Lemma 2(iii), it follows that $\frac{5-\sqrt{17}}{2} < \lim_{p\uparrow 1} \beta(p) < \frac{5}{4}$; that is, the highest bias for which a pooling equilibrium can be supported lies in the interval corresponding to β_I^M in an informed audition (see Figure 2.2a for an example). Therefore, I need to compare $\mathbb{E}[\Pi_B|\beta \in \beta_B^L]$ with $\mathbb{E}[\Pi_I|\beta \in \beta_I^L]$ and $\mathbb{E}[\Pi_I|\beta \in \beta_I^M]$ to make the above conclusion. From Proposition 1(i) and Proposition 2,

$$\begin{aligned}&\mathbb{E}[\Pi_I|\beta \in \beta_I^L] > \mathbb{E}[\Pi_B|\beta \in \beta_B^L],\\ \Rightarrow\ &\frac{p}{2}\left[\frac{2}{2+\beta}\left(1 - \frac{\beta}{2}\right) + \left(1 + \frac{\beta}{2}\right)\right] > \frac{2(1+p)+\beta}{4+2\beta}p,\end{aligned}$$

which is true for any $0 < p < 1$ and $\beta > -2$. From Proposition 1(ii) and Proposition 2,

$$\begin{aligned}&\mathbb{E}[\Pi_I|\beta \in \beta_I^M] > \mathbb{E}[\Pi_B|\beta \in \beta_B^L],\\ \Rightarrow\ &\frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2} - 1\right)\right] + \frac{p}{2}\left[\frac{2}{2+\beta}\left(1 - \frac{\beta}{2}\right) + \left(1 + \frac{\beta}{2}\right)\right] > \frac{2(1+p)+\beta}{4+2\beta}p,\end{aligned}$$

which is true for any $\frac{1}{5} < p < 1$ and $\beta > -2$. By Lemma 2, the lower bound on the prior is more restrictive to support any pooling equilibrium. By assumption, the evaluator's bias is non-negative. Therefore, both inequalities are always satisfied and the evaluator's expected net utility in a blind audition is strictly lower.

(ii) For $\beta \in \beta_I^H \cap \beta_B^H = (\frac{6}{5}, 2)$, I show that the evaluator's net utility in a blind audition is strictly greater than in an informed audition. From Proposition 1(iii) and Proposition 3,

$$\mathbb{E}[\Pi_B|\beta \in \beta_B^H] > \mathbb{E}[\Pi_I|\beta \in \beta_I^H],$$
$$\Rightarrow p\frac{4+\beta}{4+2\beta} > \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{1}{2+\beta}\left(-\frac{\beta}{2}\right) + \frac{1+\beta}{2+\beta}\left(\frac{\beta}{2}\right)\right],$$

which is true for any $-\frac{1}{8} < p \leqslant 1$ and $1 - \sqrt{8p+1} < \beta < \sqrt{8p+1} + 1$. By assumption, the lower and upper bound on the prior are more restrictive, and the evaluator's bias lies in the interval $[0, 2]$. Therefore, the above inequality is always satisfied.

(iii) For $\beta \in \beta_I^M \cap \beta_B^H = (\sqrt{17} - 3, \frac{6}{5}]$, I show that the evaluator's net utility in a blind audition is strictly greater than his expected utility in an informed audition if $0 < p < \frac{2-\beta}{2}$. From Proposition 1(ii) and Proposition 3,

$$\mathbb{E}[\Pi_B|\beta \in \beta_B^H] > \mathbb{E}[\Pi_I|\beta \in \beta_I^M],$$
$$\Rightarrow p\frac{4+\beta}{4+2\beta} > \frac{1-p}{2}\left[\frac{\beta}{2+\beta}\left(\frac{\beta}{2}-1\right)\right] + \frac{p}{2}\left[\frac{2}{2+\beta}\left(1-\frac{\beta}{2}\right) + \left(1+\frac{\beta}{2}\right)\right],$$

which is true for any $p \leqslant 1$ and $0 < \beta < 2 - 2p$.

(iv) For $\beta \in \beta_I^M \cap \beta_B^H = (\sqrt{17}-3, \frac{6}{5}]$, it follows from (iii) that the strict reverse inequality

$$\mathbb{E}[\Pi_I|\beta \in \beta_I^M] > \mathbb{E}[\Pi_B|\beta \in \beta_B^H]$$

holds for any $p \leqslant 1$ and $\beta > 2 - 2p$. □

Proof of Proposition 5. (i) For $\beta \in \beta_B^L := [0, \sqrt{p^2 + 16p} - p - 2]$ and $\frac{1}{3} \leqslant p < 1$, the applicant's expected utility is weakly greater in an blind audition:

$$U(e_H, \eta_L) \geqslant U(O, \eta_L, f),$$
$$\Rightarrow \frac{2(1+p)+\beta}{4+2\beta}w - 2 \geqslant 0.$$

By Lemma 2, the above inequality is the constraint on the range of β for which a pooling equilibrium can be supported and, therefore, always satisfied.

(ii) For $\beta \in \beta_B^H := (\sqrt{17} - 3, 2)$, the applicant does not participate in either of the auditions and her expected utility is zero, as shown in Figure 2.4a. □

Proof of Proposition 6. (i) First, note that, for $\beta \in [0, 2(1-p)]$, the applicant's expected utility is weakly greater in an informed audition:

$$U(e_H, \eta_H, f) \geqslant U(e_H, \eta_H),$$
$$\Rightarrow \frac{2}{2+\beta} \geqslant \frac{2(1+p)+\beta}{4+2\beta},$$
$$\Rightarrow \beta \leqslant 2(1-p).$$

However, because the upper bound $\beta(p) := \sqrt{p^2 + 16p} - p - 2$, for which a pooling equilibrium can be supported, may be more restrictive than the above inequality, I need to find the range of priors for which either constraint is more restrictive. The indifference condition is

$$\sqrt{p^2 + 16p} - p - 2 = 2(1-p),$$
$$\Rightarrow p^* = \frac{25 - \sqrt{561}}{2} \approx 0.65728.$$

For priors below p^*, $\beta(p)$ is the more restrictive constraint. For priors above p^*, the upper bound for which a high-ability female prefers an informed audition is given by $\beta \leqslant 2(1-p)$.

(ii) For $\beta \in \beta_B^H := (\sqrt{17} - 3, 2)$, the applicant's expected utility is strictly greater in a blind audition, as shown in Figure 2.4b. □

Proof of Proposition 7. (i) First, note that, for $\beta \in [0, \frac{5-\sqrt{17}}{2})$, the applicant does not participate in an informed audition. If $\beta \geqslant \frac{5-\sqrt{17}}{2}$, the applicant exerts low effort in an informed audition. Therefore, I need to compare

$$U(e_H, \eta_L) \geqslant U(e_L, \eta_L, m),$$
$$\Rightarrow \frac{2(1+p)+\beta}{4+2\beta} w - 2 \geqslant \frac{\beta}{2+\beta} w - \frac{1}{2},$$
$$\Rightarrow \beta \leqslant -\sqrt{p^2 + 2p + 49} + p + 7.$$

However, because the upper bound $\beta(p) := \sqrt{p^2 + 16p} - p - 2$ for which a pooling equilibrium can be supported may be more restrictive than the above inequality, I need to find the range of priors for which either constraint is more restrictive. The indifference condition is

$$\sqrt{p^2 + 16p} - p - 2 = -\sqrt{p^2 + 2p + 49} + p + 7,$$
$$\Rightarrow p^{**} = \frac{31 - 7\sqrt{17}}{4} \approx 0.53457.$$

For priors below p^{**}, $\beta(p)$ is the more restrictive constraint. For priors above p^{**}, the upper bound for which a low-ability male prefers a blind audition is given by $\beta \leqslant -\sqrt{p^2 + 2p + 49} + p + 7$.

(ii) For $\beta \in \beta_B^H := (\sqrt{17} - 3, 2)$, the applicant's expected utility is strictly greater in an informed audition, as shown in Figure 2.4c. □

Proof of Proposition 8. For $\beta \in \beta_B^L := [0, \sqrt{p^2 + 16p} - p - 2]$ and $\frac{1}{3} \leqslant p < 1$, the applicant's expected utility is strictly greater in an informed audition audition:

$$U(e_H, \eta_H, m) > U(e_H, \eta_H),$$
$$\Rightarrow 1 > \frac{2(1+p)+\beta}{4+2\beta},$$
$$\Rightarrow \beta > 2(p-1).$$

Given Lemma 2 and the assumption that evaluator's bias is non-negative, the inequality is always satisfied. For $\beta \in \beta_B^H := (\sqrt{17}-3, 2)$, the applicant's expected utility is also strictly greater in an informed audition, as shown in Figure 2.4d. $\square$

B.2 Chapter 2 Proofs

Proof of Proposition 1. I first prove the optimal pricing, discussed informally in Section 3.2.

Lemma 1 (Optimal Pricing)**.** *If the firm sells h to the whole market, then quality shoppers are indifferent between buying and not buying, and $p_h^* = p_{Q,h}$. If the firm sells h to a fraction of the market, then conscious shoppers are indifferent between buying and not buying, and $p_h^{**} = p_{C,h}$. If the firm sells m to the whole market, then all shoppers are indifferent, and $p_m^* = p_{C,m} = p_{Q,m}$.*

Proof of Lemma 1. Suppose to the contrary that Q are not indifferent when they buy product h at price p_h; that is, $V_Q(1;h) > V_Q(0;h) = 0$. This implies $p_h < p_{Q,h} = q(h, x_h) < q(h, x_h) + \theta\nu(h) = p_{C,h}$. Therefore, F could increase p_h by some amount ε without losing Q or C as customers, and F would strictly increase its profit. Hence, charging $p_h < p_{Q,h}$ can never be optimal. At $p_h^* = p_{Q,h}$, $V_Q(1;h) = V_Q(0;h) = 0$. The same logic applies in the other two cases. $\square$

The remainder of the proof proceeds in two steps: First, I derive the firm's optimal effort and maximised profit for a given choice of production technology. Second, I compare the firm's maximised profit across production technologies for different values of θ. Let $q'(x_k) := \frac{\partial q(x_k,k)}{\partial x_k}$.

First, suppose that $k = m$. Given Assumption 2, I can focus on the case that F sells m to the whole market. From Lemma 1 and Assumption 1, F sets $p_m^* = q(x_m, m) = A_m x_m$. Therefore, F's profit from choosing the effort level is $\Pi(p_m^*, x_m; m) = (C+Q)[q(x_m, m) - \frac{c_m}{2}x_m^2] = [A_m x_m - \frac{c_m}{2}x_m^2]$, which is maximised at $x_m^* = \frac{q'(x_m^*)}{c_m} = \frac{A_m}{c_m}$. Consequently, F's maximised profit is $\Pi(p_m^*, x_m^*; m) = \frac{A_m^2}{2c_m} > 0$. Therefore, given $k = m$, the firm always chooses (p_m^*, x_m^*) to sell to the whole market.

Now, fix $k = h$. There are two scenarios: (i) selling to the whole market at $p_h^* = q(x_h, h) = A_h x_h$; or (ii) selling to C only at $p_h^{**} = q(x_h, h) + \theta\nu(h) = A_h x_h + \theta$. Suppose $p_h^* = A_h x_h$. Then F's profit from choosing the effort level is $\Pi(p_h^*, x_h; h) = (C+Q)[A_h x_h - \frac{c_h}{2}x_h^2]$ which is maximised at $x_h^* = \frac{q'(x_h^*)}{c_h} = \frac{A_h}{c_h}$. Consequently, F's maximised profit is $\Pi(p_h^*, x_h^*; h) = \frac{A_h^2}{2c_h} > 0$. Suppose $p_h^{**} = A_h x_h + \theta$. Then, $x_h^{**} = x_h^*$ with $\Pi(p_h^{**}, x_h^{**}; h) =$

$C\frac{A_h^2}{2c_h} + C\theta > 0$. Therefore, given $k = h$, the firm chooses to sell to C only at p_h^{**} over the whole market at p_h^* if

$$\Pi(p_h^{**}, x_h^{**}; h) = C\Pi(p_h^*, x_h^*; h) + C\theta \geqslant \Pi(p_h^*, x_h^*; h),$$
$$\theta \geqslant \frac{1-C}{C}\Pi(p_h^*, x_h^*; h). \tag{B.1}$$

Second, suppose (B.1) does not hold: F sells to the whole market under either production process. Then F prefers to produce product m over h if $\Pi(p_m^*, x_m^*; m) > \Pi(p_h^*, x_h^*; h)$. Given $c_h > c_m > 0$, this rearranges to $\left(\frac{A_m}{A_h}\right)^2 > \frac{c_m}{c_h}$. As $\left(\frac{A_m}{A_h}\right)^2 \geqslant 1$ and $1 > \frac{c_m}{c_h} > 0$, the inequality always holds. If it is optimal to sell to the whole market, the profit is greater for $k = m$. Now, suppose (B.1) holds: F sells h to a fraction of the market. Then F prefers to produce product h over m if

$$\Pi(p_h^{**}, x_h^{**}; h) = C\Pi(p_h^*, x_h^*; h) + C\theta > \Pi(p_m^*, x_m^*; m),$$
$$\theta > \frac{1}{C}\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h). \tag{B.2}$$

Given that (B.2) is more stringent than (B.1), that is, $\Pi(p_m^*, x_m^*; m) > \Pi(p_h^*, x_h^*; h)$, the equilibrium production technology is $k = h$ whenever (B.2) holds. The condition rearranges to $C\theta > Q\Pi(p_m^*, x_m^*; m) + C\big[\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\big]$, as is stated in the proposition. □

Proof of Proposition 2. Define $\bar{p} := \bar{q} + \theta$ and $\bar{x} := \frac{\bar{q}}{h}$. First, note that the firm's profit under a binding standard is $\Pi(\bar{p}, \bar{x}; h) = C\left[\bar{q} + \theta - \frac{c_h}{2}\left(\frac{\bar{q}}{A_h}\right)^2\right] < \Pi(p_h^{**}, x_h^{**}; h)$ when selling h to C. I am interested in the conditions under which the unregulated firm chooses to sell h to C, but the regulated firm chooses to sell m to the whole market:

$$\Pi(p_h^{**}, x_h^{**}; h) > \Pi(p_m^*, x_m^*; m) > \Pi(\bar{p}, \bar{x}; h). \tag{B.3}$$

The second inequality in (B.3) rearranges as follows:

$$\Pi(p_m^*, x_m^*; m) = \frac{q(x_m^*, m)}{2} > C\left[\bar{q} + \theta - \frac{\bar{q}^2}{2q(x_h^{**}, h)}\right] = \Pi(\bar{p}, \bar{x}; h),$$
$$\bar{q}^2 - 2q(x_h^{**}, h)\bar{q} + \frac{q(x_m^*, m)q(x_h^{**}, h)}{C} - 2q(x_h^{**}, h)\theta > 0.$$

Given that $\bar{q} > q(x_h^{**}, h)$, the unique solution is

$$\bar{q} > q(x_h^{**}, h) + \sqrt{q(x_h^{**}, h)^2 - \frac{q(x_m^*, m)q(x_h^{**}, h)}{C} + 2q(x_h^{**}, h)\theta} =: \check{q}.$$

Note that the first inequality in (B.3) implies $q(x_h^{**}, h) > \frac{q(x_m^*, m)}{C} - 2\theta$. Substituting for $q(x_h^{**}, h)$ yields

$$\check{q} > q(x_h^{**}, h) + \sqrt{q(x_h^{**}, h)\left(\frac{q(x_m^*, m)}{C} - 2\theta\right) - \frac{q(x_m^*, m)q(x_h^{**}, h)}{C} + 2q(x_h^{**}, h)\theta} = q(x_h^{**}, h).$$

□

Proof of Proposition 3. First, in any such equilibrium the two firms must make the same profit. Otherwise, a firm would prefer to switch to the other production process, and try capturing the other firm's shoppers. In particular, a firm can reduce the price or increase quality. However, I show that a quality increase is more costly. Hence, the firms compete in prices.

For example, an ε price reduction on h reduces profit by $\Pi(p_h^{**}, x_h^{**}; h) - \Pi(p_h^{**} - \varepsilon, x_h^{**}; h) = C\varepsilon$, and increases C's utility by ε. An ε quality increase reduces profit by $\Pi(p_h^{**}, x_h^{**}; h) - \Pi(p_h^{**}, x_h^{**} + \varepsilon; h) = CA_h\varepsilon + C\frac{c_h}{2}\varepsilon^2$, and increases C's utility by $A_h\varepsilon$. The profit reduction from a price discount is, thus, proportional to C's utility increase. The profit reduction from a quality increase, in contrast, is more than proportional. In other words, a price discount is the more effective means: Profit is increasing linearly in price, whereas the effort to increase quality comes with a convex cost $c(x_k)$.

Second, note that, if the other firm was not taken into account, F_i would set p_h^{**} and F_j would set p_m^* so as to extract all surplus from their respective shoppers. Given that $\Pi(p_h^{**}, x_h^{**}; h) \gtrless Q\Pi(p_m^*, x_m^*; m)$ and if I assume that price adjustment is one-sided, there are two ways to equate F_i's and F_j's profit: (i) F_i lowers p_h^{**} by $\bar{\varepsilon}$ while F_j sets p_m^*; or (ii) F_j lowers p_m^* by $\hat{\varepsilon}$ while F_i sets p_h^{**}. However, (ii) cannot be an equilibrium: C are indifferent between buying and not buying h at p_h^{**}, but C are left with strictly positive utility from buying m at $p_m^* - \hat{\varepsilon} = p_{C,m} - \hat{\varepsilon}$ with $\hat{\varepsilon}$ implicitly defined by $\Pi(p_h^{**}, x_h^{**}; h) = Q\Pi(p_m^* - \hat{\varepsilon}, x_m^*; m)$. Consequently, C would want to deviate to buying m.

In case (i), both C and Q are indifferent between buying and not buying m at p_m^*, but C are left with strictly positive utility from buying h at $p_h^{**} - \bar{\varepsilon}$. Therefore, I can focus on a deviation of Q. Quality shoppers do not have an incentive to deviate to buying h if $\bar{\varepsilon}$ is sufficiently low:

$$V_Q(p_h^{**} - \bar{\varepsilon}, x_h^{**}) = V_Q(p_h^{**}, x_h^{**}) + \bar{\varepsilon} = A_h x_h^{**} - (A_h x_h^{**} + \theta) + \bar{\varepsilon} \leqslant 0,$$
$$\bar{\varepsilon} \leqslant \theta. \tag{B.4}$$

The price discount that equates profits is given by

$$\Pi(p_h^{**} - \bar{\varepsilon}, x_h^{**}; h) = C\Pi(p_h^*, x_h^*; h) + C\theta - C\bar{\varepsilon} = Q\Pi(p_m^*, x_m^*; m),$$
$$\bar{\varepsilon} = \theta - \left[\left(\frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\right]. \tag{B.5}$$

In the proposed equilibrium with a one-sided price reduction on h, F_i cannot have an incentive to deviate and reduce the price further to p_h^* such that Q switch to buying h:

$$\Pi(p_h^{**} - \bar{\varepsilon}, x_h^{**}; h) = Q\Pi(p_m^*, x_m^*; m) \geqslant \Pi(p_h^*, x_h^*; h), \tag{B.6}$$

$$\frac{A_m}{A_h} \geqslant \left[\frac{1}{1-C}\frac{c_m}{c_h}\right]^{\frac{1}{2}}.$$

Note that $\frac{A_m}{A_h} \geqslant 1$ and $\left(\frac{c_m}{c_h}\right)^{\frac{1}{2}} < 1$. However, for $C > 0$, $\left(\frac{1}{1-C}\right)^{\frac{1}{2}} > 1$. Therefore, (B.6) is not implied by the parameter restrictions but is necessary for equilibrium existence, which ensures that F_i does not have a profitable deviation. I will use (B.6) later to obtain the second condition in Proposition 3. Moreover, (B.6) implies that the second term on the right-hand side of (B.5) is strictly positive:

$$\left[\left(\frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\right] \geqslant \frac{1}{C}\Pi(p_h^*, x_h^*; h) - \Pi(p_h^*, x_h^*; h) = \frac{1-C}{C}\Pi(p_h^*, x_h^*; h) > 0.$$

Thus, substituting (B.6) into (B.5) implies that $\bar{\varepsilon} < \theta$. Given that this condition is stronger than (B.4), Q do not have an incentive to deviate. Also F_j cannot have an incentive to deviate and reduce the price to such an extent that C switch to buying m. Under pricing strategy $p_h^{**} - \bar{\varepsilon}$, C are left with strictly positive utility $\bar{\varepsilon}$.

Therefore, F_j needs to reduce its price by $\varepsilon \geqslant \bar{\varepsilon}$ to induce C to switch. To make such undercutting unprofitable for F_j, $\bar{\varepsilon}$ must be sufficiently large. In particular, setting $p_m^* - \bar{\varepsilon}$ (or less) to sell to both Q and C must lead to a weakly lower profit for F_j than selling only to Q at p_m^*:

$$\begin{aligned} Q\Pi(p_m^*, x_m^*; m) &\geqslant \Pi(p_m^* - \bar{\varepsilon}, x_m^*; m) = \Pi(p_m^*, x_m^*; m) - \bar{\varepsilon}, \\ \bar{\varepsilon} &\geqslant C\Pi(p_m^*, x_m^*; m). \end{aligned} \tag{B.7}$$

Substituting the price discount, given by (B.5), into (B.7) yields the first condition in the proposition:

$$\begin{aligned} &\theta - \left[\frac{1-C}{C}\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h)\right] \geqslant C\Pi(p_m^*, x_m^*; m), \\ &\theta \geqslant \left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h). \end{aligned} \tag{B.8}$$

Note that this first condition implies also that the price discount is strictly positive. I can substitute further for $\Pi(p_m^*, x_m^*; m)$ from (B.6) to obtain the second condition in the proposition:

$$\begin{aligned} \theta &\geqslant \left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) \\ &\geqslant \left(C + \frac{1-C}{C}\right)\left[\frac{1}{1-C}\Pi(p_h^*, x_h^*; h)\right] - \Pi(p_h^*, x_h^*; h) = \left(\frac{C}{1-C} + \frac{1-C}{C}\right)\Pi(p_h^*, x_h^*; h). \end{aligned}$$

Finally, note that the equal-profit constraint prevents F_i from selling h to C at a marginally higher price: F_i realises that this deviation results in zero profit. Suppose F_i plans to increase its price by $\epsilon \in (0, \bar{\varepsilon})$. As C enjoy a strictly positive utility, they will continue to buy h. However, F_j expects F_i to make a higher profit: It would switch the production process, and would plan to offer h to C at a price $p_h^{**} - \bar{\varepsilon} + \frac{\epsilon}{2}$. Price competition in h in this static setting implies that $p_h = c(x_h^{**})$ instantaneously. Thus, F_i is better off selling at $p_h^{**} - \bar{\varepsilon}$, and making profit $\Pi(p_h^{**} - \bar{\varepsilon}, x_h^{**}; h) > 0$. □

Proof of Proposition 4. First, the economic surplus when the two firms compete in selling m to the whole market is $ESM := (C+Q)A_m x_m^* - (C+Q)c(x_m^*) = \Pi(p_m^*, x_m^*; m)$. The specialised equilibrium is efficient if

$$ESS := Q\Pi(p_m^*, x_m^*; m) + C\Pi(p_h^*, x_h^*; h) + C\theta \geqslant \Pi(p_m^*, x_m^*; m) =: ESM,$$

$$\theta \geqslant \Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h). \tag{B.9}$$

Given that the specialised equilibrium exists, (B.8) holds and $\left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) > \Pi(p_m^*, x_m^*; m)$ for $C \in (0,1)$. Therefore, the existence of a specialised equilibrium implies (B.9). □

Proof of Proposition 5. First, F_i and F_j cannot have an incentive to merge to sell m to the whole market. Given that the firms share the monopolist's profit equally, this implies

$$\Pi(p_h^{**} - \bar{\varepsilon}, x_h^{**}; h) = Q\Pi(p_m^*, x_m^*; m) \geqslant \frac{1}{2}(C+Q)\Pi(p_m^*, x_m^*; m),$$

$$\frac{1}{2} \geqslant C.$$

Second, F_i and F_j cannot have an incentive to merge to sell h to market segment C. This implies

$$\Pi(p_h^{**} - \bar{\varepsilon}, x_h^{**}; h) = Q\Pi(p_m^*, x_m^*; m) \geqslant \frac{1}{2}\Pi(p_h^{**}, x_h^{**}; h),$$

$$Q\Pi(p_m^*, x_m^*; m) \geqslant \frac{1}{2}[C\Pi(p_h^*, x_h^*; h) + C\theta],$$

$$\frac{2(1-C)}{C}\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) \geqslant \theta. \tag{B.10}$$

□

Proof of Corollary 1. From Proposition 3, given that a specialised equilibrium exists, (B.8) holds. Moreover, from Proposition 5, F_i and F_j do not have an incentive to collude as h-monopolist if (B.10) holds. Combining these inequalities yields

$$\frac{2(1-C)}{C}\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h) \geqslant \theta \geqslant \left(C + \frac{1-C}{C}\right)\Pi(p_m^*, x_m^*; m) - \Pi(p_h^*, x_h^*; h).$$

Therefore, if

$$\left(C+\frac{1-C}{C}\right)\Pi(p_m^*,x_m^*;m)-\Pi(p_h^*,x_h^*;h)>\frac{2(1-C)}{C}\Pi(p_m^*,x_m^*;m)-\Pi(p_h^*,x_h^*;h),$$
$$C+\frac{1-C}{C}>\frac{2(1-C)}{C},$$

then merging to an h-monopolist in the specialised equilibrium is always profitable. Given that $C \in (0,1)$, the above inequality is uniquely solved by $C > \frac{\sqrt{5}-1}{2} =: \bar{C}$, where $\bar{C} > \frac{1}{2}$. From the proof of Proposition 5, collusion as m-monopolist is profitable if $C > \frac{1}{2}$. □

Proof of Proposition 6. Because $\varphi_H > \frac{1}{2} > \varphi_L$, conformist shoppers will convert into conscious shoppers if and only if the firm chooses handmade production. Therefore, the conjectured separating equilibrium exists if the firm prefers handmade production only after observing φ_H, which is equivalent to $\eta \in [\max\{0, \hat{\eta}(\varphi_H)\}, \hat{\eta}(\varphi_L)]$. Under Assumption 3,

$$\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)+\theta}<1,$$
$$[(1-\varphi_L)-(1-\varphi_H)]\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)+\theta}<(1-\varphi_L)\varphi_H-(1-\varphi_H)\varphi_L,$$
$$(1-\varphi_L)\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)}-(1-\varphi_L)\varphi_H<(1-\varphi_H)\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)}-(1-\varphi_H)\varphi_L,$$
$$\hat{\eta}(\varphi_H):=\frac{1}{1-\varphi_H}\left[\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)}-\varphi_H\right]<\frac{1}{1-\varphi_L}\left[\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)}-\varphi_L\right]=:\hat{\eta}(\varphi_L),$$

and

$$\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)+\theta}<1,$$
$$\hat{\eta}(\varphi_L):=\frac{1}{1-\varphi_L}\left[\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)+\theta}-\varphi_L\right]<1.$$

Because $\hat{\eta}(\varphi_H) < \hat{\eta}(\varphi_L) < 1$, the set is non-empty if $0 < \hat{\eta}(\varphi_L)$. This latter inequality rearranges to the condition $\theta < \frac{\Pi(p_m^*,x_m^*;m)}{\varphi_L} - \Pi(p_h^*,x_h^*;h)$ in the proposition. If follows immediately that $\theta < \Pi(p_m^*,x_m^*;m)/\varphi_H - \Pi(p_h^*,x_h^*;h)$ whenever $0 < \hat{\eta}(\varphi_H)$.

Finally, I show that my assumption $\eta \geqslant \tilde{\eta}(\varphi_H)$ is implied by $\eta \geqslant \hat{\eta}(\varphi_H)$ so that the firm that sells a handmade product does not prefer deviating to p_h^* to sell also to quality shoppers:

$$\Pi(p_h^*,x_h^*;h)<\Pi(p_m^*,x_m^*;m),$$
$$\tilde{\eta}(\varphi_H):=\frac{1}{1-\varphi_H}\left[\frac{\Pi(p_h^*,x_h^*;h)}{\Pi(p_h^*,x_h^*;h)+\theta}-\varphi_H\right]<\frac{1}{1-\varphi_H}\left[\frac{\Pi(p_m^*,x_m^*;m)}{\Pi(p_h^*,x_h^*;h)+\theta}-\varphi_H\right]=:\hat{\eta}(\varphi_H).$$

□

B.3 Chapter 3 Proofs

Proof of Lemma 1. The proof under **credulity** proceeds in two steps. We first consider a credulous player i when at least half of the preceding players have agreed with her type. We then consider a credulous player i when more than half have disagreed with θ_i.

First, consider player $i \leqslant 2M_i + 1$ and suppose that $\theta_i = 0$. Player i chooses the low action if $\lambda_i \geqslant \lambda_i(M_i, 0) = (2\mathbb{E}[\bar{s}_{-i}|M_i, 0] - 1)/(2\mathbb{E}[\bar{s}_{-i}|M_i, 0])$, where $\mathbb{E}[\bar{s}_{-i}|M_i, 0] =$

$$\Pr(\omega = 1|M_i, \theta_i = 0)\mathbb{E}[\bar{s}_{-i}|\omega = 1, M_i, \theta_i = 0] + \Pr(\omega = 0|M_i, \theta_i = 0)\mathbb{E}[\bar{s}_{-i}|\omega = 0, M_i, \theta_i = 0]. \tag{B.11}$$

To compute (B.11) when $i \leqslant 2M_i + 1$, note that $\Pr(\omega = 1|M_i, \theta_i = 0) =$

$$\frac{\pi^{i-(M_i+1)}(1-\pi)^{M_i+1}}{\pi^{i-(M_i+1)}(1-\pi)^{M_i+1} + \pi^{M_i+1}(1-\pi)^{i-(M_i+1)}} = \frac{(1-\pi)^{2(M_i+1)-i}}{\pi^{2(M_i+1)-i} + (1-\pi)^{2(M_i+1)-i}}$$

by Bayes' rule and $\mathbb{E}[\bar{s}_{-i}|\omega = 1, M_i, \theta_i = 0] =$

$$\frac{i-1-M_i}{N-1} + \frac{1}{N-1}\sum_{j=i+1}^{N} \mathbb{E}[s_j|\omega = 1, M_i, \theta_i = 0]$$
$$= \frac{i-1-M_i}{N-1} + \frac{1}{N-1}\sum_{j=i+1}^{N} \Pr(\theta_j = 1|\omega = 1) = \frac{i-1-M_i}{N-1} + \pi\left(\frac{N-i}{N-1}\right).$$

Therefore, $\mathbb{E}[\bar{s}_{-i}|M_i, 0] =$

$$\frac{(1-\pi)^{2(M_i+1)-i}}{\pi^{2(M_i+1)-i} + (1-\pi)^{2(M_i+1)-i}}\left[\frac{i-1-M_i}{N-1} + \pi\left(\frac{N-i}{N-1}\right)\right]$$
$$+ \frac{\pi^{2(M_i+1)-i}}{\pi^{2(M_i+1)-i} + (1-\pi)^{2(M_i+1)-i}}\left[\frac{i-1-M_i}{N-1} + (1-\pi)\left(\frac{N-i}{N-1}\right)\right]$$
$$= \frac{i-1-M_i}{N-1} + \frac{(1-\pi)^{2(M_i+1)-i}\pi + \pi^{2(M_i+1)-i}(1-\pi)}{\pi^{2(M_i+1)-i} + (1-\pi)^{2(M_i+1)-i}}\left(\frac{N-i}{N-1}\right). \tag{B.12}$$

Remark 1. $\partial\mathbb{E}[\bar{s}_{-i}|M_i, 0]/\partial M_i < 0$.

Proof of Remark 1. Since $\partial/\partial M_i[(i-1-M)/(N-1)] = -1/(N-1) < 0$, it is sufficient to show that the second term of (B.12) is decreasing in M_i. To ease notation, define $f(M_i) := (1-\pi)^{2(M_i+1)-i}\pi + \pi^{2(M_i+1)-i}(1-\pi)$ and $g(M_i) := \pi^{2(M_i+1)-i} + (1-\pi)^{2(M_i+1)-i}$. Then,

$$\frac{d}{dM_i}f(M_i) = 2\pi\ln(1-\pi)(1-\pi)^{2(M_i+1)-i} + 2(1-\pi)\ln(\pi)\pi^{2(M_i+1)-i} < 0$$

and

$$\frac{d}{dM_i}g(M_i) = 2\ln(\pi)\pi^{2(M_i+1)-i} + 2\ln(1-\pi)(1-\pi)^{2(M_i+1)-i} < 0.$$

Since $(N-i)/(N-1) > 0$ for $N > i$, we need to show that $g(M_i)\frac{d}{dM_i}f(M_i) - f(M_i)\frac{d}{dM_i}g(M_i) < 0$. The left-hand side of this inequality is equivalent to

$$[\pi^a + (1-\pi)^a] \cdot [2\pi \ln(1-\pi)(1-\pi)^a + 2(1-\pi)\ln(\pi)\pi^a] \\ - [\pi(1-\pi)^a + (1-\pi)\pi^a] \cdot [2\ln(\pi)\pi^a + 2\ln(1-\pi)(1-\pi)^a],$$

where $a := 2(M_i + 1) - i$. Simplifying this difference yields

$$2\pi \ln(1-\pi)[(1-\pi)\pi]^a + 2(1-\pi)\ln(\pi)\pi^{2a} + 2\pi\ln(1-\pi)(1-\pi)^{2a} + 2(1-\pi)\ln(\pi)[(1-\pi)\pi]^a \\ - \\ \big[2\pi\ln(\pi)[(1-\pi)\pi]^a + 2\pi\ln(1-\pi)(1-\pi)^{2a} + 2(1-\pi)\ln(\pi)\pi^{2a} + 2(1-\pi)\ln(1-\pi)[(1-\pi)\pi]^a\big].$$

Notice that the second and seventh terms, as well as the third and sixth, cancel out. Taking common factors of the remaining terms leads to $g(M_i)\frac{d}{dM_i}f(M_i) - f(M_i)\frac{d}{dM_i}g(M_i) =$

$$\begin{aligned} &2\ln(1-\pi)[(1-\pi)\pi]^a[\pi - (1-\pi)] + 2\ln(\pi)[(1-\pi)\pi]^a[(1-\pi) - \pi] \\ =&2[(1-\pi)\pi]^a(2\pi - 1)[\ln(1-\pi) - \ln(\pi)]. \end{aligned}$$

Since $(2\pi - 1) > 0$ for all π, we need to check that $[\ln(1-\pi) - \ln(\pi)] < 0$. Notice that at $\pi = 1/2$ this expression is equal to zero. Therefore, we need to show that the expression is strictly decreasing since $\pi > 1/2$. Differentiation yields $d/d\pi[\ln(1-\pi) - \ln(\pi)] = -1/[\pi(1-\pi)] < 0$. □

Since $i \leqslant 2M_i + 1$, we have $\mathbb{E}[\bar{s}_{-i}|M_i, 0] < \frac{1}{2}$. To see this, note that $\partial\mathbb{E}[\bar{s}_{-i}|M_i, 0]/\partial M_i < 0$ by Remark 1. Therefore, it is sufficient to show that $\mathbb{E}[\bar{s}_{-i}|M_i, 0] < 1/2$ at $M_i = (i-1)/2$. Substitution yields

$$\mathbb{E}\left[\bar{s}_{-i}\middle|\frac{i-1}{2}, 0\right] = \frac{i-1}{2(N-1)} + 2\pi(1-\pi)\frac{N-i}{N-1} < \frac{1}{2} \quad \Leftrightarrow \quad \pi(1-\pi) < \frac{1}{4},$$

which holds for all π as $\pi(1-\pi)$ is a strictly concave function with a maximum value of $1/4$ at $\pi = 1/2$. $\mathbb{E}[\bar{s}_{-i}|M_i, 0] < 1/2$ implies that $\lambda_i(M_i, 0) < 0$. As $\lambda_i \geqslant 0$ by assumption, player i always finds it optimal to choose the low action in accordance with her type; that is, $s_i^*(0) = 0$. By an analogous argument, player i always chooses the high action when $\theta_i = 1$.

Now consider player $i > 2M_i + 1$ and suppose that $\theta_i = 0$. Again, player i chooses the low action if $\lambda_i \geqslant \lambda_i(M_i, 0)$. To compute (B.11) when $i > 2M_i + 1$, note that $\Pr(\omega = 1|M_i, \theta_i = 0) =$

$$\frac{\pi^{i-(M_i+1)}(1-\pi)^{M_i+1}}{\pi^{i-(M_i+1)}(1-\pi)^{M_i+1} + \pi^{M_i+1}(1-\pi)^{i-(M_i+1)}} = \frac{\pi^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}$$

by Bayes' rule while $\mathbb{E}[\bar{s}_{-i}|\omega = 1, M_i, \theta_i = 0]$ remains as above. Therefore, $\mathbb{E}[\bar{s}_{-i}|M_i, 0] =$

$$\frac{\pi^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left[\frac{i-1-M_i}{N-1} + \pi\left(\frac{N-i}{N-1}\right)\right]$$
$$+ \frac{(1-\pi)^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left[\frac{i-1-M_i}{N-1} + (1-\pi)\left(\frac{N-i}{N-1}\right)\right]$$
$$= \frac{i-1-M_i}{N-1} + \frac{\pi^{i-1-2M_i} + (1-\pi)^{i-1-2M_i}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right). \tag{B.13}$$

Conversely, consider player $i > 2M_i + 1$ and suppose that $\theta_i = 1$. Player i chooses the high action if $\lambda_i \geqslant \lambda_i(M_i, 1) = (1 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1])/(2 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1])$, where $\mathbb{E}[\bar{s}_{-i}|M_i, 1] =$

$$\Pr(\omega = 1|M_i, \theta_i = 1)\mathbb{E}[\bar{s}_{-i}|\omega = 1, M_i, \theta_i = 1] + \Pr(\omega = 0|M_i, \theta_i = 1)\mathbb{E}[\bar{s}_{-i}|\omega = 0, M_i, \theta_i = 1]. \tag{B.14}$$

To compute (B.14) if $i > 2M_i + 1$ note that $\Pr(\omega = 1|M_i, \theta_i = 1) =$

$$\frac{\pi^{M_i+1}(1-\pi)^{i-(M_i+1)}}{\pi^{M_i+1}(1-\pi)^{i-(M_i+1)} + \pi^{i-(M_i+1)}(1-\pi)^{M_i+1}} = \frac{(1-\pi)^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}$$

by Bayes' rule and $\mathbb{E}[\bar{s}_{-i}|\omega = 1, M_i, \theta_i = 1] =$

$$\frac{M_i}{N-1} + \frac{1}{N-1}\sum_{j=i+1}^{N}\mathbb{E}[s_j|\omega = 1, M_i, \theta_i = 1] = \frac{M_i}{N-1} + \frac{1}{N-1}\sum_{j=i+1}^{N}\Pr(\theta_j = 1|\omega = 1),$$
$$= \frac{M_i}{N-1} + \pi\left(\frac{N-i}{N-1}\right).$$

Therefore, $\mathbb{E}[\bar{s}_{-i}|M_i, 1] =$

$$\frac{(1-\pi)^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left[\frac{M_i}{N-1} + \pi\left(\frac{N-i}{N-1}\right)\right]$$
$$+ \frac{\pi^{i-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left[\frac{M_i}{N-1} + (1-\pi)\left(\frac{N-i}{N-1}\right)\right]$$
$$= \frac{M_i}{N-1} + \frac{(1-\pi)^{i-2(M_i+1)}\pi + \pi^{i-2(M_i+1)}(1-\pi)}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right).$$

Finally, note that the numerator of $\lambda_i(M_i, 0)$ can be written as $2\mathbb{E}[\bar{s}_{-i}|M_i, 0] - 1 =$

$$2\left[\frac{i-(M_i+1)}{N-1} + \frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right] - 1,$$
$$= 1 + \frac{2i - 2(N-1) - 2(M_i+1)}{N-1} + 2\left[\frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right],$$
$$= 1 - \frac{2M_i}{N-1} - 2\left[1 - \frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\right]\left(\frac{N-i}{N-1}\right),$$
$$= 1 - \frac{2M_i}{N-1} - 2\left[\frac{\pi^{i-2(M_i+1)}(1-\pi) + (1-\pi)^{i-2(M_i+1)}(1-(1-\pi))}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\right]\left(\frac{N-i}{N-1}\right),$$

$$= 1 - 2\left[\frac{M_i}{N-1} + \frac{\pi^{i-2(M_i+1)}(1-\pi) + (1-\pi)^{i-2(M_i+1)}\pi}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right]$$
$$= 1 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1],$$

whilst the denominator can be written as $2\mathbb{E}[\bar{s}_{-i}|M_i, 0] =$

$$2\left[\frac{i-(M_i+1)}{N-1} + \frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right],$$
$$= 2 + \frac{2i - 2(N-1) - 2(M_i+1)}{N-1} + 2\left[\frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right],$$
$$= 2 - \frac{2M_i}{N-1} - 2\left(\frac{N-i}{N-1}\right) + 2\left[\frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\left(\frac{N-i}{N-1}\right)\right],$$
$$= 2 - \frac{2M_i}{N-1} - 2\left[1 - \frac{\pi^{i+1-2(M_i+1)} + (1-\pi)^{i+1-2(M_i+1)}}{\pi^{i-2(M_i+1)} + (1-\pi)^{i-2(M_i+1)}}\right]\left(\frac{N-i}{N-1}\right),$$
$$= 2 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1].$$

Therefore, $\lambda_i(M_i, 0) = \lambda_i(M_i, 1)$. Thus, a player with $\lambda_i < \lambda_i(M_i, \theta_i)$ chooses $s_i^*(\theta_i) = 1-\theta_i$.

Scepticism We use the fact that $\lim_{\pi\downarrow 1/2} \lambda_i(M_i, \theta_i) = \lim_{\pi\downarrow 1/2} \lambda_i(h_i^S, \theta_i)$.

We first consider player i when at least half of the preceding players have agreed with her type. Using the above fact, we proved in the previous part that $\lim_{\pi\downarrow 1/2} \lambda_i(h_i^S, \theta_i) < 0$ when $M_i \geqslant (i-1)/2$. We now consider how editing the history affects the sceptical player's cut-off when π increases.

First, suppose that $M_i \geqslant (i-2+A_i^{=}-A_i^{\neq})/2$. Then, $\partial\lambda_i(h_i^S, \theta_i)/\partial\pi \leqslant 0$ from Additional Result 1 so that $\lambda_i(h_i^S, \theta_i) < 0$ for all $\pi \in (1/2, 1)$. Hence, if $M_i \geqslant \max\{(i-1)/2, (i-2+A_i^{=}-A_i^{\neq})/2\}$, a sceptical player i never falsifies her type, proving case (i).

Now, suppose that $M_i < (i-2+A_i^{=}-A_i^{\neq})/2$. Then, $\partial\lambda_i(h_i^S, \theta_i)/\partial\pi > 0$ from Additional Result 1 so that the cut-off may become positive as π increases. In particular, $\lambda_i(h_i^S, \theta_i) = 0$ is solved by $\hat{\pi}(N)$ when $(i-1)/2 \leqslant M_i < (i-2+A_i^{=}-A_i^{\neq})/2$. Given that this threshold on the informativeness is decreasing in N, there exists a minimum group size $\hat{N}$ such that for $N > \hat{N}$, we have $\hat{\pi}(N) < 1$. The minimum group size that solves $\hat{\pi}(\hat{N}) = 1$ is $\hat{N} = 2M_i + 1$. Hence, for $N > \hat{N}$, a sceptical player i with $\lambda_i < \lambda_i(h_i^S, \theta_i)$ falsifies her type in an informative environment $\pi > \hat{\pi}(N)$. For $N \leqslant \hat{N}$, falsification never occurs, proving case (ii).

We now consider player i when less than half of the preceding players have agreed with her type. Using the above fact, we proved in the previous part that $\lim_{\pi\downarrow 1/2} \lambda_i(h_i^S, \theta_i) > 0$ when $M_i < (i-1)/2$. We again consider how editing the history affects the sceptical player's cut-off when π increases. Note here that $M_i < (i-1)/2$ is equivalent to $M_i \leqslant (i-2)/2$ given integer values.

First, suppose that $M_i \leqslant (i-2+A_i^{=}-A_i^{\neq})/2$. Then, $\partial\lambda_i(h_i^S,\theta_i)/\partial\pi \geqslant 0$ from Additional Result 1 so that $\lambda_i(h_i^S,\theta_i) > 0$ for all $\pi \in (1/2,1)$. Hence, if $M_i \leqslant \min\{(i-2)/2,(i-2+A_i^{=}-A_i^{\neq})/2\}$, a sceptical player i with $\lambda_i < \lambda_i(h_i^S,\theta_i)$ falsifies her type, proving case (iv).

Now, suppose that $M_i > (i-2+A_i^{=}-A_i^{\neq})/2$. Then, $\partial\lambda_i(h_i^S,\theta_i)/\partial\pi < 0$ from Additional Result 1 so that the cut-off may become negative as π increases. In particular, $\lambda_i(h_i^S,\theta_i) = 0$ is solved by $\check{\pi}(N)$ when $(i-2+A_i^{=}-A_i^{\neq})/2 < M_i < (i-1)/2$. Given that this threshold on the informativeness is decreasing in N, there exists a minimum group size $\check{N}$ such that for $N > \hat{N}$, we have $\hat{\pi}(N) < 1$. The minimum group size that solves $\hat{\pi}(\hat{N}) = 1$ is $\check{N} = 2(i-1-M_i)+1$. Hence, for $N > \check{N}$, a sceptical player i with $\lambda_i < \lambda_i(h_i^S,\theta_i)$ falsifies her type in an uninformative environment $\pi < \check{\pi}(N)$. For $N \leqslant \check{N}$, falsification only necessitates $\lambda_i < \lambda_i(h_i^S,\theta_i)$, proving case (iii). □

Proof of Proposition 1. We start by proving part (i). The inequality chain ensures existence of $\check{\pi}(N)$ for the minority type and $\hat{\pi}(N)$ for the majority type while $N > \hat{N}$ ensures $\{\check{\pi}(N),\hat{\pi}(N)\} < 1$. To derive the chain, suppose that type $\theta_i = 0$ is in the strict majority at history h_i when including her type. From Lemma 1, the threshold $\hat{\pi}(N)$ exists if

$$(i-1)/2 \leqslant M_i(0) < (i-2+A_i^{=}(0)-A_i^{\neq}(0))/2, \tag{B.15}$$

where $M_i(0) := \sum_{j\in h_i}\mathbb{I}\{s_j = 0\}$. From Lemma 1, the threshold $\check{\pi}(N)$ for $\theta_i = 1$ exists if

$$(i-2+A_i^{=}(1)-A_i^{\neq}(1))/2 < M_i(1) < (i-1)/2, \tag{B.16}$$

where $M_i(1) := \sum_{j\in h_i}\mathbb{I}\{s_j = 1\}$. We relate inequality chains (B.15) and (B.16) using (a) $M_i(1) = i-1-M_i(0)$, (b) $A_i^{=}(1) = A_i^{\neq}(0)$, and (c) $A_i^{\neq}(1) = A_i^{=}(0)$. To do so, we substitute conditions (a)-(c) into chain (B.16) and rearrange:

$$\begin{aligned}(i-2+A_i^{\neq}(0)-A_i^{=}(0))/2 < i-1-M_i(0) < (i-1)/2,\\ (i-1)/2 < M_i(0) < (i+A_i^{=}(0)-A_i^{\neq}(0))/2.\end{aligned} \tag{B.17}$$

The lower bound for the minority player in chain (B.17) is more stringent compared to the lower bound for the majority player in chain (B.15). On the other hand, the upper bound for the majority player is more stringent. Therefore, inequality chains (B.15) and (B.16) hold simultaneously whenever

$$(i-1)/2 < M_i(0) < (i-2+A_i^{=}(0)-A_i^{\neq}(0))/2. \tag{B.18}$$

Suppose condition (B.18) holds so that $\hat{\pi}(N)$ and $\check{\pi}(N)$ exist. For the sceptical type $\theta_i = 0$,

$$\mathbb{E}[\bar{s}_{-i}|h_i^S,0] = \frac{i-1-M_i(0)}{N-1} + \frac{\pi^{a_0+b_0+1}+(1-\pi)^{a_0+b_0+1}}{\pi^{a_0+b_0}+(1-\pi)^{a_0+b_0}}\left(\frac{N-i}{N-1}\right), \tag{B.19}$$

where $a_0 := i - 2(M_i(0)+1)$ and $b_0 := A_i^{=}(0) - A_i^{\neq}(0)$. For the sceptical type $\theta_i = 1$,

$$\mathbb{E}[\bar{s}_{-i}|h_i^S, 1] = \frac{M_i(1)}{N-1} + \frac{\pi^{1-(a_1+b_1)} + (1-\pi)^{1-(a_1+b_1)}}{\pi^{-(a_1+b_1)} + (1-\pi)^{-(a_1+b_1)}}\left(\frac{N-i}{N-1}\right), \tag{B.20}$$

where $a_1 := i - 2(M_i(1)+1)$ and $b_1 := A_i^{=}(1) - A_i^{\neq}(1)$. For a later comparison, we substitute (a) $b_1 = -b_0$, (b) $a_1 = -a_0 - 2$, and (c) $M_i(1) = i - 1 - M_i(0)$ into (B.20):

$$\mathbb{E}[\bar{s}_{-i}|h_i^S, 1] = \frac{i-1-M_i(0)}{N-1} + \frac{\pi^{a_0+b_0+3} + (1-\pi)^{a_0+b_0+3}}{\pi^{a_0+b_0+2} + (1-\pi)^{a_0+b_0+2}}\left(\frac{N-i}{N-1}\right). \tag{B.21}$$

By definition, $\hat{\pi}(N)$ and $\check{\pi}(N)$ solve $\lambda_i(h_i^S, \theta_i) = 0$, which rearranges to $\mathbb{E}[\bar{s}_{-i}|h_i^S, \theta_i] = 1/2$. Therefore, $\hat{\pi}(N)$ solves

$$\begin{aligned} \frac{1}{2} &= \frac{i-1-M_i(0)}{N-1} + \frac{\hat{\pi}^{a_0+b_0+1} + (1-\hat{\pi})^{a_0+b_0+1}}{\hat{\pi}^{a_0+b_0} + (1-\hat{\pi})^{a_0+b_0}}\left(\frac{N-i}{N-1}\right), \\ \underbrace{\frac{N+1+2(M_i(0)-i)}{2(N-i)}}_{=:Y(N)} &= \frac{\hat{\pi}^{a_0+b_0+1} + (1-\hat{\pi})^{a_0+b_0+1}}{\hat{\pi}^{a_0+b_0} + (1-\hat{\pi})^{a_0+b_0}}, \end{aligned} \tag{B.22}$$

by setting (B.19) equal to one-half. Further, $\check{\pi}(N)$ solves

$$\begin{aligned} \frac{1}{2} &= \frac{i-1-M_i(0)}{N-1} + \frac{\check{\pi}^{a_0+b_0+3} + (1-\check{\pi})^{a_0+b_0+3}}{\check{\pi}^{a_0+b_0+2} + (1-\check{\pi})^{a_0+b_0+2}}\left(\frac{N-i}{N-1}\right), \\ Y(N) &= \frac{\check{\pi}^{a_0+b_0+3} + (1-\check{\pi})^{a_0+b_0+3}}{\check{\pi}^{a_0+b_0+2} + (1-\check{\pi})^{a_0+b_0+2}}, \end{aligned} \tag{B.23}$$

by setting (B.21) equal to one-half. Moreover, $Y(\hat{N}) = Y(2M_i(0)+1) = 1$ and

$$\frac{\partial Y(N)}{\partial N} = \frac{i - 2M_i(0) - 1}{2(N-i)^2} < 0,$$

given that $2M_i(0) \geqslant i$ from the upper bound in condition (B.18). Therefore, under condition $N > \hat{N}$ of Proposition 1, we have $Y(N) < 1$.

To compare the right-hand sides of (B.22) and (B.23) for $\pi \in (1/2, 1)$, let $c_0 := a_0 + b_0$. The lower bound in condition (B.18) implies $c_0 > 0$. This restriction on the exponent, in turn, implies that

$$MAJ := \frac{\pi^{c_0+1} + (1-\pi)^{c_0+1}}{\pi^{c_0} + (1-\pi)^{c_0}} < \frac{\pi^{c_0+3} + (1-\pi)^{c_0+3}}{\pi^{c_0+2} + (1-\pi)^{c_0+2}} =: MIN$$

for all $\pi \in (1/2, 1)$. Therefore, $Y(N)$ crosses MIN at a lower π for all $N > \hat{N}$. This proves that $\check{\pi}(N) < \hat{\pi}(N)$ for all $N > \hat{N}$ (Figure B.4). Therefore, the interval $(\check{\pi}(N), \hat{\pi}(N)]$ exists under the conditions of Proposition 1.

Part (ii) applies when $\check{\pi}(N)$, but not $\hat{\pi}(N)$, exists. We need $M_i(0) \geqslant \max\{(i-1)/2, (i-2+A_i^{=}(0) - A_i^{\neq}(0))/2\}$ from Lemma 1 and (B.17), which form the inequality chain in part (ii):

$$\max\{(i-1)/2 + \varepsilon, (i-2+A_i^{=}(0) - A_i^{\neq}(0))/2\} \leqslant M_i(0) < (i + A_i^{=}(0) - A_i^{\neq}(0))/2. \tag{B.24}$$

Suppose condition (B.24) holds. Then condition $N > \hat{N}$ of Proposition 1 ensures that $\breve{\pi}(N) < 1$ and existence of interval $(\breve{\pi}(N), 1)$ follows.

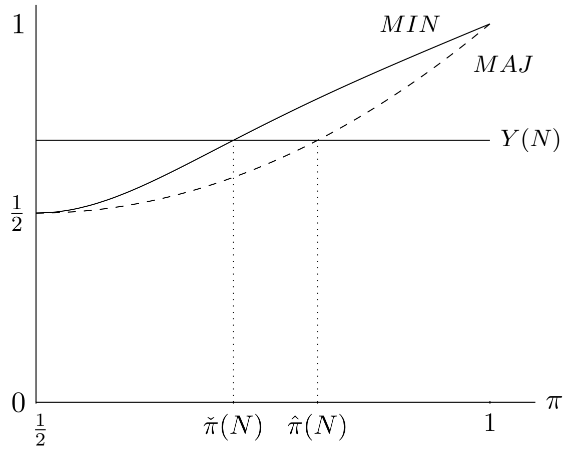


Figure B.4: $N > \hat{N}$.

□

Proof of Lemma 2. We solve the model under rationality with $N = 4$ backwards. In deriving the equilibrium strategies of player 2 and 3, we assume $s_1^*(\theta_1) = \theta_1$. In the last step, we show that player 1 does not have an incentive to deviate from this strategy. For ease of exposition, we set $\theta_1 = 0$.

In the last period, player 4 observes one of the following histories: $h_{40} = \{0, 0, 0\}$, $h_{42} = \{0, 1, 0\}$, $h_{43} = \{0, 0, 1\}$, or $h_{44} = \{0, 1, 1\}$. The corresponding cut-offs for $\theta_4 \in \{0, 1\}$ are

$$\begin{aligned} &\lambda_4^R(h_{40}, 1) = \frac{1}{2} &&\lambda_4^R(h_{42}, 1) = \lambda_4^R(h_{43}, 1) = \frac{1}{4} &&\lambda_4^R(h_{44}, 1) < 0, \\ &\lambda_4^R(h_{40}, 0) = -\frac{1}{0} &&\lambda_4^R(h_{42}, 0) = \lambda_4^R(h_{43}, 0) < 0 &&\lambda_4^R(h_{44}, 0) = \frac{1}{4}. \end{aligned}$$

Hence, we need to consider **three cases** for player 3: $\lambda_4 < 1/4$, $1/4 \leqslant \lambda_4 < 1/2$, and $1/2 \leqslant \lambda_4$.

Case I Assume first $\lambda_4 < 1/4$. Consequently, player 3 infers that player 4 will agree with the majority of her predecessors. Player 3 also takes into account whether player 2 revealed θ_2 through her action.

(i) Suppose player 3 infers that the low action at $t = 2$ could come from either $\theta_2 = 1$ or $\theta_2 = 0$. Player 3 then discards player 2's action in calculating $\Pr(\omega = 1 | h_{30}, \theta_3)$. That is, player 3 expects $h_{30} = \{0, 0\}$ to involve $s_2^*(\theta_2) = 0$ but $h_{32} = \{0, 1\}$ to involve $s_2^*(\theta_2) = \theta_2$.

Consider first $\theta_3 = 1$. At history $h_{30} = \{0,0\}$, $\Pr(\omega = 1|h_{30}, 1) = 1/2$. If player 3 takes the high action, player 4 observes $h_{43} = \{0,0,1\}$. Given $\lambda_4 < 1/4$, player 4 will always take the low action. Conversely, if player 3 takes the low action, player 4 observes $h_{40} = \{0,0,0\}$. Again, player 4 will always take the low action. Therefore, player 3 expects utility

$$\mathbb{E}[U(1;1,\lambda_3)|h_{30}] = -(1-\lambda_3)\left(1 - \frac{0+0+0}{3}\right)^2 - \lambda_3(1-1)^2 = -(1-\lambda_3),$$
$$\mathbb{E}[U(0;1,\lambda_3)|h_{30}] = -(1-\lambda_3)\left(0 - \frac{0+0+0}{3}\right)^2 - \lambda_3(0-1)^2 = -\lambda_3,$$

respectively, so that $s_3^*(1, h_{30}) = 0$ if $\lambda_3 < 1/2 =: \lambda_3^R(h_{30}, 1)$. At history $h_{32} = \{0,1\}$, $\Pr(\omega = 1|h_{32}, 1) = \frac{\pi^2(1-\pi)}{\pi^2(1-\pi)+(1-\pi)^2\pi}$. If player 3 takes the high action, player 4 observes $h_{44} = \{0,1,1\}$. Given $\lambda_4 < 1/4$, player 4 will always take the high action. Conversely, if player 3 takes the low action, player 4 observes $h_{42} = \{0,1,0\}$ and will always take the low action. Therefore, player 3 expects utility

$$\mathbb{E}[U(1;1,\lambda_3)|h_{32}] = -(1-\lambda_3)\left(1 - \frac{0+1+1}{3}\right)^2 - \lambda_3(1-1)^2 = -\frac{1-\lambda_3}{9},$$
$$\mathbb{E}[U(0;1,\lambda_3)|h_{32}] = -(1-\lambda_3)\left(0 - \frac{0+1+0}{3}\right)^2 - \lambda_3(0-1)^2 = -\frac{1-\lambda_3}{9} - \lambda_3,$$

respectively, so that $s_3^*(1, h_{32}) = 1$.

Consider now $\theta_3 = 0$. At history $h_{30} = \{0,0\}$, $\Pr(\omega = 1|h_{30}, 0) = \frac{(1-\pi)^2}{(1-\pi)^2+\pi^2}$. If player 3 takes the high action, $\mathbb{E}[U(1;0,\lambda_3)|h_{30}] = -(1-\lambda_3) - \lambda_3$. Conversely, if player 3 takes the low action, $\mathbb{E}[U(0;0,\lambda_3)|h_{30}] = 0$. Therefore, $s_3^*(0, h_{30}) = 0$. The strategy is ideal: player 3 with $\theta_3 = 0$ at h_{30} does not face a trade-off between matching her type and matching the average action. At history h_{32}, $\Pr(\omega = 1|h_{32}, 0) = \frac{\pi(1-\pi)^2}{\pi(1-\pi)^2+(1-\pi)\pi^2}$. If player 3 takes the high action, $\mathbb{E}[U(1;0,\lambda_3)|h_{32}] = -(1-\lambda_3)/9 - \lambda_3$. Conversely, if player 3 takes the low action, $\mathbb{E}[U(0;0,\lambda_3)|h_{32}] = -(1-\lambda_3)/9$. Therefore, $s_3^*(0, h_{32}) = 0$. In sum, given $\lambda_4 < 1/4$, $s_3^*(\theta_3, h_{30}) = 0$ if $\lambda_3 < 1/2$ whereas $s_3^*(\theta_3, h_{32}) = \theta_3$.

(ii) Suppose player 3 infers that the low action at $t = 2$ comes from $\theta_2 = 0$. Player 3 then does not discard player 2's action in calculating $\Pr(\omega = 1|h_{30}, \theta_3)$. However, as player 4 always agrees with the majority, and player 3 cannot affect the majority at h_{30}, $\mathbb{E}[U(1;1,\lambda_3)|h_{30}]$ and $\mathbb{E}[U(0;1,\lambda_3)|h_{30}]$ remain the same. Hence, the change from $s_2^*(\theta_2) = 0$ to $s_2^*(\theta_2) = \theta_2$ at h_{30} does not change $\lambda_3^R(h_{30}, 1)$. It remains unchanged that $s_3^*(0, h_{30}) = 0$.

Case II Assume now $1/4 \leqslant \lambda_4 < 1/2$.

(i) Suppose player 3 infers that the low action at $t = 2$ could come from either type. Consider first $\theta_3 = 1$. If player 3 takes the high action at h_{30}, player 4 observes $h_{43} = \{0, 0, 1\}$. Given $\lambda_4 \geqslant 1/4$, player 4 will always reveal truthfully. Conversely, if player 3 takes the low action, player 4 observes $h_{40} = \{0, 0, 0\}$. Given $\lambda_4 < 1/2$, player 4 will always take the low action. Therefore,

$$\mathbb{E}[U(1; 1, \lambda_3)|h_{30}] = -\frac{1-\lambda_3}{2}\left(1 - \frac{0+0+1}{3}\right)^2 - \frac{1-\lambda_3}{2}\left(1 - \frac{0+0+0}{3}\right)^2 = -\frac{13(1-\lambda_3)}{18},$$

and $\mathbb{E}[U(0; 1, \lambda_3)|h_{30}] = -\lambda_3$. Hence, $s_3^*(1, h_{30}) = 0$ if $\lambda_3 < 13/31 =: \lambda_3^R(h_{30}, 1)$. If player 3 takes the high action at h_{32}, player 4 observes $h_{44} = \{0, 1, 1\}$. Given $\lambda_4 \geqslant 1/4$, player 4 will always reveal truthfully. If player 3 takes the low action, player 4 observes $h_{42} = \{0, 1, 0\}$. Again, she will always reveal truthfully. Therefore, player 3 expects utility

$$\mathbb{E}[U(1; 1, \lambda_3)|h_{32}] = \frac{u}{v}\left[-(1-\lambda_3)\left(1 - \frac{0+1+1}{3}\right)^2\right] + \frac{w}{v}\left[-(1-\lambda_3)\left(1 - \frac{0+1+0}{3}\right)^2\right],$$

$$\begin{aligned}\mathbb{E}[U(0; 1, \lambda_3)|h_{32}] = \frac{u}{v}&\left[-(1-\lambda_3)\left(0 - \frac{0+1+1}{3}\right)^2\right] \\ &+ \frac{w}{v}\left[-(1-\lambda_3)\left(0 - \frac{0+1+0}{3}\right)^2\right] - \lambda_3,\end{aligned}$$

where $u := \pi^3(1-\pi) + (1-\pi)^3\pi$, $v := \pi^2(1-\pi) + (1-\pi)^2\pi$, and $w := 2\pi^2(1-\pi)^2$. Given that $u > w$, $\mathbb{E}[U(0; 1, \lambda_3)|h_{32}] + \lambda_3$ is lower than $\mathbb{E}[U(1; 1, \lambda_3)|h_{32}]$. Hence, $\mathbb{E}[U(0; 1, \lambda_3)|h_{32}]$ is also lower, and $s_3^*(1, h_{32}) = 1$. Intuitively, because player 3 does not alter player 4's behaviour and has evidence in favour of state $\omega = 1$, she maximises the probability of matching her action to her successor's action by revealing truthfully. Simultaneously, she avoids disutility from falsification.

Consider now $\theta_3 = 0$. If player 3 takes the high action at h_{30}, $\mathbb{E}[U(1; 0, \lambda_3)|h_{30}] = -13(1-\lambda_3)/18 - \lambda_3$. Conversely, if player 3 takes the low action, $\mathbb{E}[U(0; 0, \lambda_3)|h_{30}] = 0$. Therefore, $s_3^*(0, h_{30}) = 0$. The strategy is ideal. If player 3 takes the high action at h_{32}, $\mathbb{E}[U(1; 0, \lambda_3)|h_{32}] = \mathbb{E}[U(0; 1, \lambda_3)|h_{32}]$. Conversely, if player 3 takes the low action, $\mathbb{E}[U(0; 0, \lambda_3)|h_{32}] = \mathbb{E}[U(1; 1, \lambda_3)|h_{32}]$. Therefore, given $u > w$, $s_3^*(0, h_{32}) = 0$. In sum, given $1/4 \leqslant \lambda_4 < 1/2$, $s_3^*(\theta_3, h_{30}) = 0$ if $\lambda_3 < 13/31$ whereas $s_3^*(\theta_3, h_{32}) = \theta_3$.

(ii) When player 3 infers that the low action at $t = 2$ comes from $\theta_2 = 0$, $\Pr(\omega = 1|h_{30}, 1) = 1 - \pi$. Therefore, $\mathbb{E}[U(1; 1, \lambda_3)|h_{30}] =$

$$-2\pi(1-\pi)(1-\lambda_3)\left(1 - \frac{0+0+1}{3}\right)^2 - [\pi^2 + (1-\pi)^2](1-\lambda_3)\left(1 - \frac{0+0+0}{3}\right)^2$$

$$= -\left[1 - \frac{10\pi(1-\pi)}{9}\right](1-\lambda_3),$$

and $\mathbb{E}[U(0;1,\lambda_3)|h_{30}] = -\lambda_3$. Hence, $s_3^*(1,h_{30}) = 0$ if $\lambda_3 < (9-10\pi(1-\pi))/(18-10\pi(1-\pi)) =: \lambda_3^R(h_{30},1)$. It remains unchanged that $s_3^*(0,h_{30}) = 0$.

Case III Assume $1/2 \leqslant \lambda_4$.

(i) Suppose player 3 knows that the low action at $t=2$ could come from either type. Consider first $\theta_3 = 1$. As player 4 reveals truthfully at any history, $\mathbb{E}[U(1;1,\lambda_3)|h_{30}] = -13(1-\lambda_3)/18$ and $\mathbb{E}[U(0;1,\lambda_3)|h_{30}] =$

$$-\frac{1-\lambda_3}{2}\left(0 - \frac{0+0+1}{3}\right)^2 - \frac{1-\lambda_3}{2}\left(0 - \frac{0+0+0}{3}\right)^2 - \lambda_3 = -\frac{1-\lambda_3}{18} - \lambda_3.$$

Therefore, $s_3^*(1,h_{30}) = 0$ if $\lambda_3 < 2/5 =: \lambda_3^R(h_{30},1)$. At h_{32}, the expected utilities are the same as when $1/4 \leqslant \lambda_4 < 1/2$. Therefore, $s_3^*(1,h_{32}) = 1$. Consider now $\theta_3 = 0$. If player 3 takes the high action at h_{30}, $\mathbb{E}[U(1;0,\lambda_3)|h_{30}] = -13(1-\lambda_3)/18 - \lambda_3$. Conversely, if player 3 takes the low action, $\mathbb{E}[U(0;0,\lambda_3)|h_{30}] = -(1-\lambda_3)/18$. Therefore, $s_3^*(0,h_{30}) = 0$. As above, $s_3^*(0,h_{32}) = 0$. In sum, given $1/2 \leqslant \lambda_4$, $s_3^*(\theta_3,h_{30}) = 0$ if $\lambda_3 < 2/5$ whereas $s_3^*(\theta_3,h_{32}) = \theta_3$.

(ii) When player 3 knows that the low action at $t=2$ comes from $\theta_2 = 0$, $\mathbb{E}[U(1;1,\lambda_3)|h_{30}] = -[1-10/9\pi(1-\pi)](1-\lambda_3)$ while $\mathbb{E}[U(0;1,\lambda_3)|h_{30}] =$

$$-2\pi(1-\pi)(1-\lambda_3)\left(0-\frac{0+0+1}{3}\right)^2 - [\pi^2+(1-\pi)^2](1-\lambda_3)\left(0-\frac{0+0+0}{3}\right)^2 - \lambda_3$$
$$= -\frac{2\pi(1-\pi)(1-\lambda_3)}{9} - \lambda_3.$$

Hence, $s_3^*(1,h_{30}) = 0$ if $\lambda_3 < (3-4\pi(1-\pi))/(6-4\pi(1-\pi))$. It remains unchanged that $s_3^*(0,h_{30}) = 0$.

Comparison for Player 3 We compare player 3's cut-off under rationality to its counterpart under credulity. The latter is given by $\lambda_3(0,1) = \frac{1-2\mathbb{E}[\bar{s}_{-3}|0,1]}{2-2\mathbb{E}[\bar{s}_{-3}|0,1]} = \frac{3-4\pi(1-\pi)}{6-4\pi(1-\pi)}$, where

$$\mathbb{E}[\bar{s}_{-3}|0,1] = \frac{(1-\pi)^{3-2}\pi + \pi^{3-2}(1-\pi)}{(1-\pi)^{3-2}+\pi^{3-2}}\left(\frac{4-3}{4-1}\right) = \frac{2(1-\pi)\pi}{3}.$$

The cut-offs coincide if player 3 correctly infers the other players to be truthful. If player 4 instead is inferred to agree with the majority ($\lambda_4 < 1/4$) or with a consensus ($1/4 \leqslant \lambda_4 < 1/2$), the rational cut-off is higher.

As the credulous cut-off is strictly increasing in π, it is bounded below by $\lim_{\pi\to 1/2}\lambda_3(0,1) = 2/5$, and above by $\lim_{\pi\to 1}\lambda_3(0,1) = 1/2$. Therefore, if $s_2^*(\theta_2) = \theta_1$ and $\lambda_4 < 1/4$ ($1/2 \leqslant \lambda_4$), the rational cut-off is higher (lower). If $s_2^*(\theta_2) = \theta_1$ and $1/4 \leqslant \lambda_4 < 1/2$, the rational cut-off is higher whenever $\pi < \frac{1}{2} + \frac{1}{2\sqrt{6}} =: \tilde{\pi}$.

	$\lambda_4 < 1/4$	$1/4 \leqslant \lambda_4 < 1/2$	$1/2 \leqslant \lambda_4$
(i) $s_2^*(\theta_2) = \theta_1$	$\frac{1}{2}$	$\frac{13}{31}$	$\frac{2}{5}$
(ii) $s_2^*(\theta_2) = \theta_2$	$\frac{1}{2}$	$\frac{9-10\pi(1-\pi)}{18-10\pi(1-\pi)}$	$\frac{3-4\pi(1-\pi)}{6-4\pi(1-\pi)}$

Table B.1: Player 3's cut-off $\lambda_3^R(h_{30},1)$ under rationality.

For player 2, we need to consider the following **six cases**.

Case I(a) Suppose $\lambda_4 < 1/4$ and $\lambda_3 < 1/2$. In the second period, player 2 observes $h_{20} = \{0\}$. Consider first $\theta_2 = 1$. At h_{20}, $\Pr(\omega = 1|h_{20},1) = 1/2$. If player 2 takes the high action, player 3 will always reveal truthfully, and player 4 will agree with player 3. Conversely, if player 2 takes the low action, player 3 and 4 will always take the low action. Hence,

$$\mathbb{E}[U(1;1,\lambda_2)|h_{20}] = -\frac{1-\lambda_2}{2}\left(1-\frac{0+1+1}{3}\right)^2 - \frac{1-\lambda_2}{2}\left(1-\frac{0+0+0}{3}\right)^2 = -\frac{5(1-\lambda_2)}{9},$$
$$\mathbb{E}[U(0;1,\lambda_2)|h_{20}] = -(1-\lambda_2)\left(0-\frac{0+0+0}{3}\right)^2 - \lambda_2(0-1)^2 = -\lambda_2,$$

respectively, so that $s_2^*(1,h_{20}) = 0$ if $\lambda_2 < 5/14 =: \lambda_2^R(h_{20},1)$. Consider now $\theta_2 = 0$. At history h_{20}, $\Pr(\omega = 1|h_{20},0) = \frac{(1-\pi)^2}{(1-\pi)^2+\pi^2}$. Given that $\mathbb{E}[U(0;0,\lambda_2)|h_{20}] = 0 > \mathbb{E}[U(1;0,\lambda_2)|h_{20}]$, $s_2^*(0,h_{20}) = 0$.

Case I(b) Suppose $\lambda_4 < 1/4$ and $\lambda_3 \geqslant 1/2$, and consider first $\theta_2 = 1$. Now, if player 2 takes the low action, player 3 will always reveal truthfully, and player 4 will always take the low action. Therefore, player 2 expects utility $\mathbb{E}[U(1;1,\lambda_2)|h_{20}] = -\frac{5(1-\lambda_2)}{9}$ and $\mathbb{E}[U(0;1,\lambda_2)|h_{20}] =$

$$-\frac{1-\lambda_2}{2}\left(0-\frac{0+1+0}{3}\right)^2 - \frac{1-\lambda_2}{2}\left(0-\frac{0+0+0}{3}\right)^2 - \lambda_2(0-1)^2 = -\frac{1-\lambda_2}{18} - \lambda_2.$$

Hence, $s_2^*(1,h_{20}) = 0$ if $\lambda_2 < 1/3 =: \lambda_2^R(h_{20},1)$. Consider now $\theta_2 = 0$. Given that this type of player 2 can induce player 4 to always agree with her by taking the low action, and simultaneously avoid the disutility from falsification, $s_2^*(0,h_{20}) = 0$.

Case II(a) Suppose $1/4 \leqslant \lambda_4 < 1/2$ and $\lambda_3 < \lambda_3^R(h_{30}, 1)$, and consider $\theta_2 = 1$. If player 2 takes the high action, player 3 and 4 will always reveal truthfully. Conversely, if player 2 takes the low action, player 3 and 4 will always take the low action. Therefore,

$$\begin{aligned}\mathbb{E}[U(1;1,\lambda_2)|h_{20}] &= -\frac{\pi^2+(1-\pi)^2}{2}(1-\lambda_2)\left(1-\frac{0+1+1}{3}\right)^2 \\ &\quad - \pi(1-\pi)(1-\lambda_2)\left(1-\frac{0+1+0}{3}\right)^2 - (1-\pi)\pi(1-\lambda_2)\left(1-\frac{0+0+1}{3}\right)^2 \\ &\quad - \frac{(1-\pi)^2+\pi^2}{2}(1-\lambda_2)\left(1-\frac{0+0+0}{3}\right)^2 = -\frac{[5-2\pi(1-\pi)](1-\lambda_2)}{9},\end{aligned}$$

and $\mathbb{E}[U(0;1,\lambda_2)|h_{20}] = -\lambda_2$. Hence, $s_2^*(1,h_{20}) = 0$ if $\lambda_2 < \frac{5-2\pi(1-\pi)}{14-2\pi(1-\pi)} =: \lambda_2^R(h_{20},1) < 5/14$. Consider now $\theta_2 = 0$. Given that $\mathbb{E}[U(0;0,\lambda_2)|h_{20}] = 0 > \mathbb{E}[U(1;0,\lambda_2)|h_{20}]$, $s_2^*(0,h_{20}) = 0$.

Case II(b) Suppose $1/4 \leqslant \lambda_4 < 1/2$ and $\lambda_3 \geqslant \lambda_3^R(h_{30}, 1)$ and consider $\theta_2 = 1$. Now, if player 2 takes the low action, player 3 will always reveal truthfully, and player 4 will take the low action if $s_3 = 0$. Therefore, $\mathbb{E}[U(1;1,\lambda_2)|h_{20}] = -\frac{[5-2\pi(1-\pi)](1-\lambda_2)}{9}$ and $\mathbb{E}[U(0;1,\lambda_2)|h_{20}] =$

$$\begin{aligned}&-\frac{\pi^2+(1-\pi)^2}{2}(1-\lambda_2)\left(0-\frac{0+1+1}{3}\right)^2 - \pi(1-\pi)(1-\lambda_2)\left(0-\frac{0+1+0}{3}\right)^2 \\ &\quad - \frac{(1-\pi)\pi+(1-\pi)^2+\pi(1-\pi)+\pi^2}{2}(1-\lambda_2)\left(0-\frac{0+0+0}{3}\right)^2 - \lambda_2 \\ &\quad = -\frac{[2-3\pi(1-\pi)](1-\lambda_2)}{9} - \lambda_2.\end{aligned}$$

Hence, $s_2^*(1,h_{20}) = 0$ if $\lambda_2 < \frac{3+\pi(1-\pi)}{12+\pi(1-\pi)} =: \lambda_2^R(h_{20},1)$. Consider now $\theta_2 = 0$. Given that this type of player 2 can induce player 4 to agree with her more often by taking the low action, and simultaneously avoid the disutility from falsification, $s_2^*(0,h_{20}) = 0$.

Case III(a) Suppose $1/2 \leqslant \lambda_4$ and $\lambda_3 < \lambda_3^R(h_{30}, 1)$, and consider $\theta_2 = 1$. If player 2 takes the high action, player 3 and 4 will always reveal truthfully. Conversely, if player 2 takes the low action, player 3 will always take the low action, and player 4 will always reveal truthfully. Therefore, player 2 expects utility $\mathbb{E}[U(1;1,\lambda_2)|h_{20}] = -\frac{[5-2\pi(1-\pi)](1-\lambda_2)}{9}$ and $\mathbb{E}[U(0;1,\lambda_2)|h_{20}] =$

$$-\frac{1-\lambda_2}{2}\left(0-\frac{0+0+1}{3}\right)^2 - \frac{1-\lambda_2}{2}\left(0-\frac{0+0+0}{3}\right)^2 - \lambda_2 = -\frac{1-\lambda_2}{18} - \lambda_2.$$

Therefore, $s_2^*(1,h_{20}) = 0$ if $\lambda_2 < \frac{9-4\pi(1-\pi)}{27-4\pi(1-\pi)} =: \lambda_2^R(h_{20},1)$. Consider now $\theta_2 = 0$. Given that this type of player 2 can induce player 3 to always agree with her by taking the low action, and simultaneously avoid the disutility from falsification, $s_2^*(0,h_{20}) = 0$.

Case III(b) Suppose $1/2 \leqslant \lambda_4$ and $\lambda_3 \geqslant \lambda_3^R(h_{30}, 1)$, and consider $\theta_2 = 1$. Now, if player 2 takes the low action, player 3 and 4 will always reveal truthfully. Therefore, player 2 expects utility $\mathbb{E}[U(1; 1, \lambda_2)|h_{20}] = -\frac{[5-2\pi(1-\pi)](1-\lambda_2)}{9}$ and

$$\begin{aligned}\mathbb{E}[U(0; 1, \lambda_2)|h_{20}] = &-\frac{\pi^2 + (1-\pi)^2}{2}(1-\lambda_2)\left(0 - \frac{0+1+1}{3}\right)^2 \\ &- \pi(1-\pi)(1-\lambda_2)\left(0 - \frac{0+1+0}{3}\right)^2 - (1-\pi)\pi(1-\lambda_2)\left(0 - \frac{0+0+1}{3}\right)^2 \\ &- \frac{(1-\pi)^2 + \pi^2}{2}(1-\lambda_2)\left(0 - \frac{0+0+0}{3}\right)^2 - \lambda_2 = -\frac{[2 - 2\pi(1-\pi)](1-\lambda_2)}{9} - \lambda_2.\end{aligned}$$

Therefore, $s_2^*(1, h_{20}) = 0$ if $\lambda_2 < \frac{1}{4} =: \lambda_2^R(h_{20}, 1)$. Because $s_1^*(\theta_1) = \theta_1$ and successors are always expected to be truthful, $\lambda_2^R(h_{20}, 1) = \lambda_2(0, 1)$. Consider now $\theta_2 = 0$. Given that this type of player 2 can maximise the probability of matching player 3 and 4's action by taking the low action, and simultaneously avoid the disutility from falsification, $s_2^*(0, h_{20}) = 0$.

Comparison for Player 2 We compare player 2's cut-off under rationality to its counterpart $\lambda_2(0, 1) = 1/N = 1/4$ under credulity. For all $\pi \in (1/2, 1)$, the lowest cut-off under rationality obtains in **Case III(b)**. Consequently, as long as the other players do not always tell the truth, the cut-off in the rational model is strictly greater. The same insight obtained for player 3 conditional on $s_2^*(\theta_2) = \theta_2$.

	$\lambda_4 < 1/4$	$1/4 \leqslant \lambda_4 < 1/2$	$1/2 \leqslant \lambda_4$
(a) $\lambda_3 < \lambda_3^R(h_{30}, 1)$	$\frac{5}{14}$	$\frac{5-2\pi(1-\pi)}{14-2\pi(1-\pi)}$	$\frac{9-4\pi(1-\pi)}{27-4\pi(1-\pi)}$
(b) $\lambda_3 \geqslant \lambda_3^R(h_{30}, 1)$	$\frac{1}{3}$	$\frac{3+\pi(1-\pi)}{12+\pi(1-\pi)}$	$\frac{1}{4}$

Table B.2: Player 2's cut-off $\lambda_2^R(h_{20}, 1)$ under rationality.

In equilibrium, players 2, 3 and 4 expect player 1 to be truthful. Therefore, we show in this last step that player 1 does not have an incentive to deviate from $s_1^*(0) = 0$ in period 1 when a deviation induces players 2, 3 and 4 to believe that player 1 has a high type. The expectation that player 1 is truthful, for example, implies that if a player 2 with $\theta_2 = 1$ is willing to falsify at equilibrium history $h_{20} = \{0\}$, then a player 2 with $\theta_2 = 0$ is willing to falsify at deviation history $\bar{h}_{20} = \{1\}$. Likewise, if a player 3 with $\theta_3 = 1$ is willing to falsify at equilibrium history $h_{30} = \{0, 0\}$, then a player 3 with $\theta_3 = 0$ is willing to falsify at deviation history $\bar{h}_{30} = \{1, 1\}$. Define the remaining deviation histories $\bar{h}_{32} = \{1, 0\}$, $\bar{h}_{40} = \{1, 1, 1\}$, $\bar{h}_{42} = \{1, 0, 1\}$, $\bar{h}_{43} = \{1, 1, 0\}$, and $\bar{h}_{44} = \{1, 0, 0\}$. We need to consider **twelve combinations** of successors' cut-offs.

Case 1 Suppose $\lambda_2 < \frac{5}{14}$, $\lambda_3 < \frac{1}{2}$ and $\lambda_4 < \frac{1}{4}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, 0, 0\}$. When $s_1(0) = 1$, she infers $\{s_2, s_3, s_4\} = \{1, 1, 1\}$. Hence, player 1's expected utilities are $\mathbb{E}[U(0; 0, \lambda_1)] = 0$ and $\mathbb{E}[U(1; 0, \lambda_1)] = -\lambda_1$ which implies that player 1 finds it optimal to truthfully reveal: $s_1^*(0) = 0$.

Case 2 Suppose $\lambda_2 \geqslant \frac{5}{14}$, $\lambda_3 < \frac{1}{2}$ and $\lambda_4 < \frac{1}{4}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 0 \text{ if } h_{30} \text{ and } \theta_3 \text{ if } h_{32}, 1 \text{ if } h_{44} \text{ and } 0 \text{ otherwise}\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] = (1-\pi)\mathbb{E}[U(0; 0, \lambda_1)|\omega = 1] + \pi\mathbb{E}[U(0; 0, \lambda_1)|\omega = 0] =$

$$-(1-\pi)(1-\lambda_1)\left[\pi^2\left(0-\frac{1+1+1}{3}\right)^2+\pi(1-\pi)\left(0-\frac{1+0+0}{3}\right)^2+(1-\pi)\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[(1-\pi)^2\left(0-\frac{1+1+1}{3}\right)^2+(1-\pi)\pi\left(0-\frac{1+0+0}{3}\right)^2+\pi\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$= -\pi(1-\pi)(1-\lambda_1) - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 1 \text{ if } \bar{h}_{30} \text{ and } \theta_3 \text{ if } \bar{h}_{32}, 0 \text{ if } \bar{h}_{44} \text{ and } 1 \text{ otherwise}\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[(1-\pi)^2\left(1-\frac{0+0+0}{3}\right)^2+(1-\pi)\pi\left(1-\frac{0+1+1}{3}\right)^2+\pi\left(1-\frac{1+1+1}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[\pi^2\left(1-\frac{0+0+0}{3}\right)^2+\pi(1-\pi)\left(1-\frac{0+1+1}{3}\right)^2+(1-\pi)\left(1-\frac{1+1+1}{3}\right)^2\right]-\lambda_1$$
$$= -[(1-\pi)^3+\pi^3](1-\lambda_1) - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

As $(1-\pi)^3 + \pi^3 > \pi(1-\pi)$ for any π, $s_1^*(0) = 0$. Player 1 avoids the disutility from falsification, and reduces the probability of $s_1 \neq \theta_2 = \theta_3$ from $(1-\pi)^3 + \pi^3$ to $\pi(1-\pi)$, thereby reducing the expected disutility from a situation in which all successors disagree with player 1.

Case 3 Suppose $\lambda_2 < \frac{1}{3}$, $\lambda_3 \geqslant \frac{1}{2}$ and $\lambda_4 < \frac{1}{4}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, \theta_3, 0\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[\pi\left(0-\frac{0+1+0}{3}\right)^2+(1-\pi)\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[(1-\pi)\left(0-\frac{0+1+0}{3}\right)^2+\pi\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$= -2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{1, \theta_3, 1\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$- (1-\pi)(1-\lambda_1)\left[(1-\pi)\left(1-\frac{1+0+1}{3}\right)^2 + \pi\left(1-\frac{1+1+1}{3}\right)^2\right]$$
$$- \pi(1-\lambda_1)\left[\pi\left(1-\frac{1+0+1}{3}\right)^2 (1-\pi)\left(1-\frac{1+1+1}{3}\right)^2\right] - \lambda_1$$
$$= -[1-2\pi(1-\pi)]\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

As $2\pi(1-\pi) < 1/2$ for any π, $\mathbb{E}[U(0;0,\lambda_1)] > \mathbb{E}[U(1;0,\lambda_1)]$ and $s_1^*(0) = 0$. Again, player 1 avoids the disutility from falsification and reduces the probability of $s_1 \neq \theta_2$ from $1-2\pi(1-\pi)$ to $2\pi(1-\pi)$.

Case 4 Suppose $\lambda_2 \geqslant \frac{1}{3}$, $\lambda_3 \geqslant \frac{1}{2}$ and $\lambda_4 < \frac{1}{4}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, \theta_3, 1 \text{ if } h_{44} \text{ and } 0 \text{ otherwise}\}$. She expects $\mathbb{E}[U(0;0,\lambda_1)] =$

$$- (1-\pi)(1-\lambda_1)\left[\pi^2\left(0-\frac{1+1+1}{3}\right)^2 + \pi(1-\pi)\left(0-\frac{1+0+0}{3}\right)^2\right.$$
$$\left. + (1-\pi)\pi\left(0-\frac{0+1+0}{3}\right)^2 + (1-\pi)^2\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$- \pi(1-\lambda_1)\left[(1-\pi)^2\left(0-\frac{1+1+1}{3}\right)^2 + (1-\pi)\pi\left(0-\frac{1+0+0}{3}\right)^2\right.$$
$$\left. + \pi(1-\pi)\left(0-\frac{0+1+0}{3}\right)^2 + \pi^2\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$= -\pi(1-\pi)(1-\lambda_1) - 2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, \theta_3, 0 \text{ if } \bar{h}_{44} \text{ and } 1 \text{ otherwise}\}$. She expects $\mathbb{E}[U(1;0,\lambda_1)] =$

$$- (1-\pi)(1-\lambda_1)\left[(1-\pi)^2\left(1-\frac{0+0+0}{3}\right)^2 + (1-\pi)\pi\left(1-\frac{0+1+1}{3}\right)^2\right.$$
$$\left. + \pi(1-\pi)\left(1-\frac{1+0+1}{3}\right)^2 + \pi^2\left(1-\frac{1+1+1}{3}\right)^2\right]$$
$$- \pi(1-\lambda_1)\left[\pi^2\left(1-\frac{0+0+0}{3}\right)^2 + \pi(1-\pi)\left(1-\frac{0+1+1}{3}\right)^2\right.$$
$$\left. + (1-\pi)\pi\left(1-\frac{1+0+1}{3}\right)^2 + (1-\pi)^2\left(1-\frac{1+1+1}{3}\right)^2\right] - \lambda_1$$
$$= -[(1-\pi)^3 + \pi^3](1-\lambda_1) - 2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

As in **Case 2**, $(1-\pi)^3 + \pi^3 > \pi(1-\pi)$ implies $s_1^*(0) = 0$. Being truthful does not alter the probability of $\theta_2 \neq \theta_3$ so that player 4 agrees with whatever player 1 announces. However, it allows player 1 to avoid the disutility from falsification, and reduce the probability of $s_1 \neq \theta_2 = \theta_3$ from $(1-\pi)^3 + \pi^3$ to $\pi(1-\pi)$ in which case player 4 would instead agree with the other two players.

Case 5 Suppose $\lambda_2 < \frac{5-2\pi(1-\pi)}{14-2\pi(1-\pi)}$, $\lambda_3 < \frac{13}{31}$ and $\frac{1}{4} \leqslant \lambda_4 < \frac{1}{2}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, 0, 0\}$. When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{1, 1, 1\}$. As in **Case 1**, $s_1^*(0) = 0$.

Case 6 Suppose $\lambda_2 \geqslant \frac{5-2\pi(1-\pi)}{14-2\pi(1-\pi)}$, $\lambda_3 < \frac{9-10\pi(1-\pi)}{18-10\pi(1-\pi)}$ and $\frac{1}{4} \leqslant \lambda_4 < \frac{1}{2}$. Thus, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 0 \text{ if } h_{30} \text{ and } \theta_3 \text{ if } h_{32}, 0 \text{ if } h_{40} \text{ and } \theta_4 \text{ otherwise}\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$
\begin{aligned}
&- (1-\pi)(1-\lambda_1)\Bigg[\pi^3\left(0 - \frac{1+1+1}{3}\right)^2 + \pi^2(1-\pi)\left(0 - \frac{1+1+0}{3}\right)^2 \\
&+ \pi(1-\pi)\pi\left(0 - \frac{1+0+1}{3}\right)^2 + \pi(1-\pi)^2\left(0 - \frac{1+0+0}{3}\right)^2 + (1-\pi)\left(0 - \frac{0+0+0}{3}\right)^2\Bigg] \\
&- \pi(1-\lambda_1)\Bigg[(1-\pi)^3\left(0 - \frac{1+1+1}{3}\right)^2 + (1-\pi)^2\pi\left(0 - \frac{1+1+0}{3}\right)^2 \\
&+ (1-\pi)\pi(1-\pi)\left(0 - \frac{1+0+1}{3}\right)^2 + (1-\pi)\pi^2\left(0 - \frac{1+0+0}{3}\right)^2 + \pi\left(0 - \frac{0+0+0}{3}\right)^2\Bigg] \\
&= -[(1-\pi)\pi^3 + \pi(1-\pi)^3](1-\lambda_1) - 4\pi^2(1-\pi)^2\left[\frac{4(1-\lambda_1)}{9}\right] - [\pi(1-\pi)^3 + (1-\pi)\pi^3]\left(\frac{1-\lambda_1}{9}\right).
\end{aligned}
$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 1 \text{ if } \bar{h}_{30} \text{ and } \theta_3 \text{ if } \bar{h}_{32}, 1 \text{ if } \bar{h}_{40} \text{ and } \theta_4$ otherwise$\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$
\begin{aligned}
&- (1-\pi)(1-\lambda_1)\Bigg[(1-\pi)^3\left(1 - \frac{0+0+0}{3}\right)^2 + (1-\pi)^2\pi\left(1 - \frac{0+0+1}{3}\right)^2 \\
&+ (1-\pi)\pi(1-\pi)\left(1 - \frac{0+1+0}{3}\right)^2 + (1-\pi)\pi^2\left(1 - \frac{0+1+1}{3}\right)^2 + \pi\left(1 - \frac{1+1+1}{3}\right)^2\Bigg] \\
&- \pi(1-\lambda_1)\Bigg[\pi^3\left(1 - \frac{0+0+0}{3}\right)^2 + \pi^2(1-\pi)\left(1 - \frac{0+0+1}{3}\right)^2 + \pi(1-\pi)\pi\left(1 - \frac{0+1+0}{3}\right)^2 \\
&+ \pi(1-\pi)^2\left(1 - \frac{0+1+1}{3}\right)^2 + (1-\pi)\left(1 - \frac{1+1+1}{3}\right)^2\Bigg] - \lambda_1 \\
&= -[(1-\pi)^4 + \pi^4](1-\lambda_1) - 2[(1-\pi)^3\pi + (1-\pi)\pi^3]\left[\frac{4(1-\lambda_1)}{9}\right] - 2\pi^2(1-\pi)^2\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.
\end{aligned}
$$

The first two terms in $\mathbb{E}[U(0; 0, \lambda_1)]$ are less negative. However, the third term is more negative than in $\mathbb{E}[U(1; 0, \lambda_1)]$. Hence, to show that $\mathbb{E}[U(0; 0, \lambda_1)] > \mathbb{E}[U(1; 0, \lambda_1)]$, we combine the second and the third term in each expected utility, cancel common factors and compare:

$$
\begin{gathered}
-4\pi^2(1-\pi)^2 4 - [\pi(1-\pi)^3 + (1-\pi)\pi^3] > -2[(1-\pi)^3\pi + (1-\pi)\pi^3]4 - 2\pi^2(1-\pi)^2, \\
-16\pi(1-\pi) - [(1-\pi)^2 + \pi^2] > -8[(1-\pi)^2 + \pi^2] - 2\pi(1-\pi), \\
[(1-\pi)^2 + \pi^2] > 2\pi(1-\pi).
\end{gathered}
$$

Given that the above holds for any π, the fall in disutility from reducing the probability of players 2 and 3, or 2 and 4 disagreeing with player 1 outweighs the rise in disutility from increasing the probability of only player 2 disagreeing with player 1. Hence, $s_1^*(0) = 0$.

Case 7 Suppose $\lambda_2 < \frac{3+\pi(1-\pi)}{12+\pi(1-\pi)}$, $\lambda_3 \geqslant \frac{13}{31}$ and $\frac{1}{4} \leqslant \lambda_4 < \frac{1}{2}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, \theta_3, 0 \text{ if } h_{40} \text{ and } \theta_4 \text{ otherwise}\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[\pi^2\left(0-\frac{0+1+1}{3}\right)^2+\pi(1-\pi)\left(0-\frac{0+1+0}{3}\right)^2+(1-\pi)\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[(1-\pi)^2\left(0-\frac{0+1+1}{3}\right)^2+(1-\pi)\pi\left(0-\frac{0+1+0}{3}\right)^2+\pi\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$= -\pi(1-\pi)\left[\frac{4(1-\lambda_1)}{9}\right] - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{1, \theta_3, 1 \text{ if } \bar{h}_{40} \text{ and } \theta_4 \text{ otherwise}\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[(1-\pi)^2\left(1-\frac{1+0+0}{3}\right)^2+(1-\pi)\pi\left(1-\frac{1+0+1}{3}\right)^2+\pi\left(1-\frac{1+1+1}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[\pi^2\left(1-\frac{1+0+0}{3}\right)^2+\pi(1-\pi)\left(1-\frac{1+0+1}{3}\right)^2+(1-\pi)\left(1-\frac{1+1+1}{3}\right)^2\right]-\lambda_1$$
$$= -[(1-\pi)^3+\pi^3]\left[\frac{4(1-\lambda_1)}{9}\right] - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

As in **Case 2 and 4**, $(1-\pi)^3 + \pi^3 > \pi(1-\pi)$ implies $s_1^*(0) = 0$. Being truthful does not alter the probability of player 2 and 4 agreeing but player 3 disagreeing with player 1. However, it allows player 1 to avoid the disutility from falsification, and reduce the probability of $s_1 \neq \theta_3 = \theta_4$ from $(1-\pi)^3 + \pi^3$ to $\pi(1-\pi)$.

Case 8 Suppose $\lambda_2 \geqslant \frac{3+\pi(1-\pi)}{12+\pi(1-\pi)}$, $\lambda_3 \geqslant \frac{9-10\pi(1-\pi)}{18-10\pi(1-\pi)}$ and $\frac{1}{4} \leqslant \lambda_4 < \frac{1}{2}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, \theta_3, 0 \text{ if } h_{40} \text{ and } \theta_4 \text{ otherwise}\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[\pi^3\left(0-\frac{1+1+1}{3}\right)^2+\pi^2(1-\pi)\left(0-\frac{1+1+0}{3}\right)^2\right.$$
$$+\pi(1-\pi)\pi\left(0-\frac{1+0+1}{3}\right)^2+\pi(1-\pi)^2\left(0-\frac{1+0+0}{3}\right)^2$$
$$\left.+(1-\pi)\pi^2\left(0-\frac{0+1+1}{3}\right)^2+(1-\pi)\pi(1-\pi)\left(0-\frac{0+1+0}{3}\right)^2+(1-\pi)^2\left(0-\frac{0+0+0}{3}\right)^2\right]$$
$$-\pi(1-\lambda_1)\left[(1-\pi)^3\left(0-\frac{1+1+1}{3}\right)^2+(1-\pi)^2\pi\left(0-\frac{1+1+0}{3}\right)^2\right.$$

$$+ (1-\pi)\pi(1-\pi)\left(0 - \frac{1+0+1}{3}\right)^2 + (1-\pi)\pi^2\left(0 - \frac{1+0+0}{3}\right)^2$$

$$+ \pi(1-\pi)^2\left(0 - \frac{0+1+1}{3}\right)^2 + \pi(1-\pi)\pi\left(0 - \frac{0+1+0}{3}\right)^2 + \pi^2\left(0 - \frac{0+0+0}{3}\right)^2\Bigg]$$

$$= -[(1-\pi)\pi^3+\pi(1-\pi)^3](1-\lambda_1)-6\pi^2(1-\pi)^2\left[\frac{4(1-\lambda_1)}{9}\right]-2[\pi(1-\pi)^3+\pi^3(1-\pi)]\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, \theta_3, 1 \text{ if } \bar{h}_{40} \text{ and } \theta_4 \text{ otherwise}\}$. She expects $\mathbb{E}[U(1;0,\lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\Bigg[(1-\pi)^3\left(1 - \frac{0+0+0}{3}\right)^2 + (1-\pi)^2\pi\left(1 - \frac{0+0+1}{3}\right)^2$$

$$+(1-\pi)\pi(1-\pi)\left(1-\frac{0+1+0}{3}\right)^2+(1-\pi)\pi^2\left(1-\frac{0+1+1}{3}\right)^2+\pi(1-\pi)^2\left(1-\frac{1+0+0}{3}\right)^2$$

$$+ \pi(1-\pi)\pi\left(1 - \frac{1+0+1}{3}\right)^2 + \pi^2\left(1 - \frac{1+1+1}{3}\right)^2\Bigg]$$

$$-\pi(1-\lambda_1)\Bigg[\pi^3\left(1-\frac{0+0+0}{3}\right)^2+\pi^2(1-\pi)\left(1-\frac{0+0+1}{3}\right)^2+\pi(1-\pi)\pi\left(1-\frac{0+1+0}{3}\right)^2$$

$$+ \pi(1-\pi)^2\left(1 - \frac{0+1+1}{3}\right)^2 + (1-\pi)\pi^2\left(1 - \frac{1+0+0}{3}\right)^2$$

$$+ (1-\pi)\pi(1-\pi)\left(1 - \frac{1+0+1}{3}\right)^2 + (1-\pi)^2\left(1 - \frac{1+1+1}{3}\right)^2\Bigg] - \lambda_1$$

$$= -[(1-\pi)^4+\pi^4](1-\lambda_1)-3[(1-\pi)^3\pi+(1-\pi)\pi^3]\left[\frac{4(1-\lambda_1)}{9}\right]-4\pi^2(1-\pi)^2\left(\frac{1-\lambda_1}{9}\right)-\lambda_1.$$

Following the same argument as in **Case 6**, $s_1^*(0) = 0$.

Case 9 Suppose $\lambda_2 < \frac{9-4\pi(1-\pi)}{27-4\pi(1-\pi)}$, $\lambda_3 < \frac{2}{5}$ and $\lambda_4 \geqslant \frac{1}{2}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, 0, \theta_4\}$. She expects $\mathbb{E}[U(0;0,\lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[\pi\left(0 - \frac{0+0+1}{3}\right)^2 + (1-\pi)\left(0 - \frac{0+0+0}{3}\right)^2\right]$$

$$-\pi(1-\lambda_1)\left[+(1-\pi)\left(0 - \frac{0+0+1}{3}\right)^2 + \pi\left(0 - \frac{0+0+0}{3}\right)^2\right]$$

$$= -2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{1, 1, \theta_4\}$. She expects $\mathbb{E}[U(1;0,\lambda_1)] =$

$$-(1-\pi)(1-\lambda_1)\left[(1-\pi)\left(1 - \frac{1+1+0}{3}\right)^2 + \pi\left(1 - \frac{1+1+1}{3}\right)^2\right]$$

$$-\pi(1-\lambda_1)\left[\pi\left(1 - \frac{1+1+0}{3}\right)^2 + (1-\pi)\left(1 - \frac{1+1+1}{3}\right)^2\right] - \lambda_1$$

$$= -[1 - 2\pi(1-\pi)]\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

Given that utilities are the same as in **Case 3**, $s_1^*(0) = 0$.

Case 10 Suppose $\lambda_2 \geqslant \frac{9-4\pi(1-\pi)}{27-4\pi(1-\pi)}$, $\lambda_3 < \frac{3-4\pi(1-\pi)}{6-4\pi(1-\pi)}$ and $\lambda_4 \geqslant \frac{1}{2}$. Therefore, when $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 0 \text{ if } h_{30} \text{ and } \theta_3 \text{ if } h_{32}, \theta_4\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$\begin{aligned}
&-(1-\pi)(1-\lambda_1)\left[\pi^3\left(0-\frac{1+1+1}{3}\right)^2 + \pi^2(1-\pi)\left(0-\frac{1+1+0}{3}\right)^2\right.\\
&\quad + \pi(1-\pi)\pi\left(0-\frac{1+0+1}{3}\right)^2 + \pi(1-\pi)^2\left(0-\frac{1+0+0}{3}\right)^2\\
&\quad \left.+ (1-\pi)\pi\left(0-\frac{0+0+1}{3}\right)^2 + (1-\pi)^2\left(0-\frac{0+0+0}{3}\right)^2\right]\\
&-\pi(1-\lambda_1)\left[(1-\pi)^3\left(0-\frac{1+1+1}{3}\right)^2 + (1-\pi)^2\pi\left(0-\frac{1+1+0}{3}\right)^2\right.\\
&\quad + (1-\pi)\pi(1-\pi)\left(0-\frac{1+0+1}{3}\right)^2 + \pi^2(1-\pi)\left(0-\frac{1+0+0}{3}\right)^2\\
&\quad \left.+ \pi(1-\pi)\left(0-\frac{0+0+1}{3}\right)^2 + \pi^2\left(0-\frac{0+0+0}{3}\right)^2\right]\\
&= -[(1-\pi)\pi^3 + \pi(1-\pi)^3](1-\lambda_1) - 4\pi^2(1-\pi)^2\left[\frac{4(1-\lambda_1)}{9}\right]\\
&\quad - [\pi(1-\pi)^3 + \pi^3(1-\pi)]\left(\frac{1-\lambda_1}{9}\right) - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).
\end{aligned}$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, 1 \text{ if } \bar{h}_{30} \text{ and } \theta_3 \text{ if } \bar{h}_{32}, \theta_4\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$\begin{aligned}
&-(1-\pi)(1-\lambda_1)\left[(1-\pi)^3\left(1-\frac{0+0+0}{3}\right)^2 + (1-\pi)^2\pi\left(1-\frac{0+0+1}{3}\right)^2\right.\\
&\quad + (1-\pi)\pi(1-\pi)\left(1-\frac{0+1+0}{3}\right)^2 + (1-\pi)\pi^2\left(1-\frac{0+1+1}{3}\right)^2\\
&\quad \left.+ \pi(1-\pi)\left(1-\frac{1+1+0}{3}\right)^2 + \pi^2\left(1-\frac{1+1+1}{3}\right)^2\right]\\
&-\pi(1-\lambda_1)\left[\pi^3\left(1-\frac{0+0+0}{3}\right)^2 + \pi^2(1-\pi)\left(0-\frac{0+0+1}{3}\right)^2\right.\\
&\quad + \pi(1-\pi)\pi\left(1-\frac{0+1+0}{3}\right)^2 + (1-\pi)^2\pi\left(0-\frac{0+1+1}{3}\right)^2\\
&\quad \left.+ (1-\pi)\pi\left(1-\frac{1+1+0}{3}\right)^2 + (1-\pi)^2\left(1-\frac{1+1+1}{3}\right)^2\right] - \lambda_1\\
&= -[(1-\pi)^4 + \pi^4](1-\lambda_1) - 2[(1-\pi)^3\pi + (1-\pi)\pi^3]\left[\frac{4(1-\lambda_1)}{9}\right]
\end{aligned}$$

$$- 2\pi^2(1-\pi)^2\left(\frac{1-\lambda_1}{9}\right) - \pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.$$

Given that the fourth term is constant across utilities, we use the same argument as in **Case 6 and 8** to conclude that $s_1^*(0) = 0$.

Case 11 Suppose $\lambda_2 < 1/4$, $\lambda_3 \geqslant \frac{2}{5}$ and $\lambda_4 \geqslant \frac{1}{2}$. $s_1(0) = 0$, player 1 infers $\{s_2, s_3, s_4\} = \{0, \theta_3, \theta_4\}$. She expects $\mathbb{E}[U(0; 0, \lambda_1)] =$

$$\begin{aligned}
&- (1-\pi)(1-\lambda_1)\Bigg[\pi^2\left(0 - \frac{0+1+1}{3}\right)^2 + \pi(1-\pi)\left(0 - \frac{0+1+0}{3}\right)^2 \\
&\qquad + (1-\pi)\pi\left(0 - \frac{0+0+1}{3}\right)^2 + (1-\pi)^2\left(0 - \frac{0+0+0}{3}\right)^2\Bigg] \\
&\qquad - \pi(1-\lambda_1)\Bigg[(1-\pi)^2\left(0 - \frac{0+1+1}{3}\right)^2 + (1-\pi)\pi\left(0 - \frac{0+1+0}{3}\right)^2 \\
&\qquad + \pi(1-\pi)\left(0 - \frac{0+0+1}{3}\right)^2 + \pi^2\left(0 - \frac{0+0+0}{3}\right)^2\Bigg] \\
&\qquad = -\pi(1-\pi)\left[\frac{4(1-\lambda_1)}{9}\right] - 2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right).
\end{aligned}$$

When $s_1(0) = 1$, player 1 infers $\{s_2, s_3, s_4\} = \{1, \theta_3, \theta_4\}$. She expects $\mathbb{E}[U(1; 0, \lambda_1)] =$

$$\begin{aligned}
&- (1-\pi)(1-\lambda_1)\Bigg[(1-\pi)^2\left(1 - \frac{1+0+0}{3}\right)^2 + (1-\pi)\pi\left(1 - \frac{1+0+1}{3}\right)^2 \\
&\qquad + \pi(1-\pi)\left(1 - \frac{1+1+0}{3}\right)^2 + \pi^2\left(1 - \frac{1+1+1}{3}\right)^2\Bigg] \\
&\qquad - \pi(1-\lambda_1)\Bigg[\pi^2\left(1 - \frac{1+0+0}{3}\right)^2 + \pi(1-\pi)\left(1 - \frac{1+0+1}{3}\right)^2 \\
&\qquad + (1-\pi)\pi\left(1 - \frac{1+1+0}{3}\right)^2 + (1-\pi)^2\left(1 - \frac{1+1+1}{3}\right)^2\Bigg] - \lambda_1 \\
&\qquad = -[(1-\pi)^3 + \pi^3]\left[\frac{4(1-\lambda_1)}{9}\right] - 2\pi(1-\pi)\left(\frac{1-\lambda_1}{9}\right) - \lambda_1.
\end{aligned}$$

As in **Case 2, 4 and 7**, $(1-\pi)^3 + \pi^3 > \pi(1-\pi)$ implies $s_1^*(0) = 0$. Being truthful does not alter the probability of either player 3 or 4 disagreeing with player 1. However, it allows player 1 to avoid the disutility from falsification, and reduce the probability of $s_1 \neq \theta_3 = \theta_4$ from $(1-\pi)^3 + \pi^3$ to $\pi(1-\pi)$.

Case 12 Suppose $\lambda_2 < 1/4$, $\lambda_3 \geqslant \frac{3-4\pi(1-\pi)}{6-4\pi(1-\pi)}$ and $\lambda_4 \geqslant \frac{1}{2}$. Under either action, player 1 infers $\{s_2, s_3, s_4\} = \{\theta_2, \theta_3, \theta_4\}$. Hence, this case has the same solution as under credulity: $s_1^*(0) = 0$. □

Proof of Theorem 1. From the proof of Lemma 2, there are two possible combinations of conformity motives that generate the mappings in cases $(i) - (iii)$ under rationality.

For case (i), these are $\lambda_2 < 5/14$, $\lambda_3 < 1/2$, $\lambda_4 < 1/4$ and $\lambda_2 < [5 - 2\pi(1-\pi)]/[14 - 2\pi(1-\pi)]$, $\lambda_3 < 13/31$, $1/4 \leqslant \lambda_4 < 1/2$. Under the assumption that $\lambda := \lambda_i = \lambda_j$ for all $i \neq j \in \mathcal{N}$, we only need to consider the lowest conformity motive in each combination: $\lambda < 1/4$ and $\lambda < \lambda_2^R(h_{20}, \theta_2) = [5 - 2\pi(1-\pi)]/[14 - 2\pi(1-\pi)]$, respectively. Given that the latter is greater, we know that $(\theta, \theta_2, \theta_3, \theta_4) \mapsto (\theta, \theta, \theta, \theta)$ whenever $\lambda < \lambda_2^R(h_{20}, \theta_2)$.

For case (ii), the combinations are $\lambda_2 \geqslant 5/14$, $\lambda_3 < 1/2$, $\lambda_4 < 1/4$ and $\lambda_2 \geqslant [5 - 2\pi(1-\pi)]/[14 - 2\pi(1-\pi)]$, $\lambda_3 < [9 - 10\pi(1-\pi)]/[18 - 10\pi(1-\pi)]$, $1/4 \leqslant \lambda_4 < 1/2$. Since $\lambda \geqslant 5/14$ and $\lambda < 1/4$ cannot hold simultaneously, the first combination cannot generate the mappings in case (ii). The second combination generates the mappings when $\lambda_2^R(h_{20}, \theta_2) \leqslant \lambda < \lambda_3^R(h_{30}, \theta_3) = [9 - 10\pi(1-\pi)]/[18 - 10\pi(1-\pi)]$.

For case (iii), the combinations are $\lambda_2 \geqslant 5/14$, $\lambda_3 \geqslant 1/2$, $\lambda_4 < 1/4$ and $\lambda_2 \geqslant \lambda_2^R(h_{20}, \theta_2)$, $\lambda_3 \geqslant \lambda_3^R(h_{30}, \theta_3)$ and $1/4 \leqslant \lambda_4 < 1/2$. Similarly, only the second combination can generate the mappings in case (iii) when $\lambda_3^R(h_{30}, \theta_3) \leqslant \lambda < 1/2$.

In the credulous model, the cut-offs follow from the proof of Lemma 1: $\lambda_2(0, \theta_2) = 1/4$, $\lambda_3(0, \theta_3) = [3 - 4\pi(1-\pi)]/[6 - 4\pi(1-\pi)]$, $\lambda_4(0, \theta_4) = 1/2$ and $\lambda_4(1, \theta_4) = 1/4$. Given that $\lambda_2(0, \theta_2) < \lambda_3(0, \theta_3) < \lambda_4(0, \theta_4)$, we know that for $\lambda < \lambda_2(0, \theta_2)$ all players agree with the first. For $\lambda_2(0, \theta_2) \leqslant \lambda < \lambda_3(0, \theta_3)$, the second player is truthful. Finally, for $\lambda_3(0, \theta_3) \leqslant \lambda < \lambda_4(0, \theta_4)$, the third player is truthful.

In the sceptical model, the reasoning is similar. However, player 3 edits the history when $M_3 = 0$ so that $h_3 = \{\theta, \theta\}$ reduces to $h_3^S = \{\theta\}$. This means that, compared to the credulous model, only $\lambda_3(0, \theta_3)$ changes and falls to $\lambda_3(h_3^S, \theta_3) = 2/5$. □

Proof of Corollary 1. In case (i) of Theorem 1, the probability of $(s_1, s_2, s_3, s_4) = (\theta, \theta, \theta, \theta)$ is unity. In (ii), this probability is $\Pr((\theta_1, \theta_2) = (\theta, \theta)) = \pi^2 + (1-\pi)^2 = 1 - 2\pi(1-\pi)$; in (iii), $\Pr((\theta_1, \theta_2, \theta_3) = (\theta, \theta, \theta)) = \pi^3 + (1-\pi)^3$; and, in (iv), $\Pr((\theta_1, \theta_2, \theta_3, \theta_4) = (\theta, \theta, \theta, \theta)) = \pi^4 + (1-\pi)^4$.

Therefore, if $\lambda \sim F[0, 1]$, the ex-ante probability of $(s_1, s_2, s_3, s_4) = (\theta, \theta, \theta, \theta)$ across $(i) - (iv)$ is

$$C := 2\pi(1-\pi) \cdot F\left(\frac{1}{4}\right) + \pi(1-\pi) \cdot F\left(\frac{3 - 4\pi(1-\pi)}{6 - 4\pi(1-\pi)}\right) + \delta,$$

$$S := 2\pi(1-\pi) \cdot F\left(\frac{1}{4}\right) + \pi(1-\pi) \cdot F\left(\frac{2}{5}\right) + \delta, \text{ and}$$

$$R := 2\pi(1-\pi) \cdot F\left(\frac{5 - 2\pi(1-\pi)}{14 - 2\pi(1-\pi)}\right) + \pi(1-\pi) \cdot F\left(\frac{9 - 10\pi(1-\pi)}{18 - 10\pi(1-\pi)}\right) + \delta,$$

in the credulous, sceptical and rational model, respectively, where $\delta := F(1/2)\pi(1-\pi)(\pi^2 + (1-\pi)^2) + \pi^4 + (1-\pi)^4$. So the effect of a rationality boost is positive, $\Delta R := R - C > 0$, whereas the effect of a scepticism boost is negative, $\Delta S := S - C < 0$. □

Proof of Corollary 2. The proof proceeds in two steps. We first calculate the expected welfare for each of the four mappings in Theorem 1. We then compare the expected welfare and derive conditions on the common conformity motive λ such that a certain mapping maximises expected welfare.

To calculate the expected welfare for a mapping, we derive the ex-post welfare for each of the 16 possible type realisations. As shown in Example B.3, the ex-post welfare in mapping (i) is either 0, $-\lambda$, -2λ or -3λ. The weight attached to each ex-ante is found by summing the probabilities in the second column of Example B.3. The probabilities, in turn, are derived using the fact that $\Pr(\omega = 1) = \Pr(\omega = 0) = 1/2$ and $\Pr(\theta_i = 0|\omega = 0) = \Pr(\theta_i = 1|\omega = 1) := \pi \in (1/2, 1)$. This gives an expected welfare of

$$\begin{aligned}\mathbb{E}[W] &= 3[(1-\pi)^3\pi + \pi^3(1-\pi)](-\lambda) + 6(1-\pi)^2\pi^2(-2\lambda) + [(1-\pi)^3\pi + \pi^3(1-\pi)](-3\lambda)\\ &= -6\pi(1-\pi)\lambda =: W_{(i)}.\end{aligned}$$

The ex-post welfare in mapping (iv) is either 0, $-(1-\lambda)4/3$ or $-(1-\lambda)16/9$. Summing the probabilities in the second column to derive their weights, we obtain

$$\begin{aligned}\mathbb{E}[W] &= 4[(1-\pi)^3\pi + \pi^3(1-\pi)]\Big(-\frac{(1-\lambda)4}{3}\Big) + 6(1-\pi)^2\pi^2\Big(-\frac{(1-\lambda)16}{9}\Big)\\ &= -\frac{16}{3}\pi(1-\pi)(1-\lambda) =: W_{(iv)}.\end{aligned}$$

As shown in Example B.3, the ex-post welfare in mappings (ii) and (iii) correspond either to (i) or (iv). Hence, for mapping (ii),

$$\begin{aligned}\mathbb{E}[W] &= 2[(1-\pi)^3\pi + \pi^3(1-\pi)](-\lambda) + 2(1-\pi)^2\pi^2(-2\lambda)\\ &\quad + 2[(1-\pi)^3\pi + \pi^3(1-\pi)]\Big(-\frac{(1-\lambda)4}{3}\Big) + 4(1-\pi)^2\pi^2\Big(-\frac{(1-\lambda)16}{9}\Big)\\ &= -2\pi(1-\pi)\lambda - \frac{8}{3}\pi(1-\pi)\Big(1+\frac{2}{3}\pi(1-\pi)\Big)(1-\lambda) =: W_{(ii)};\end{aligned}$$

for mapping (iii),

$$\begin{aligned}\mathbb{E}[W] &= [(1-\pi)^3\pi + \pi^3(1-\pi)](-\lambda)\\ &\quad + 3[(1-\pi)^3\pi + \pi^3(1-\pi)]\Big(-\frac{(1-\lambda)4}{3}\Big) + 6(1-\pi)^2\pi^2\Big(-\frac{(1-\lambda)16}{9}\Big)\\ &= -\pi(1-\pi)[\pi^2 + (1-\pi)^2]\lambda - 4\pi(1-\pi)\Big(1+\frac{2}{3}\pi(1-\pi)\Big)(1-\lambda) =: W_{(iii)}.\end{aligned}$$

Finally, we compare the expected welfare of the four mappings to find the welfare-maximising mapping from the utilitarian policy maker's perspective:

$$W^* = \max\left\{W_{(i)}, W_{(ii)}, W_{(iii)}, W_{(iv)}\right\} = \begin{cases} W_{(i)}, & \text{for } 0 \leqslant \lambda < \lambda', \\ W_{(ii)}, & \text{for } \lambda' < \lambda < \lambda'', \\ W_{(iii)}, & \text{for } \lambda'' < \lambda < \lambda''', \\ W_{(iv)}, & \text{for } \lambda''' < \lambda \leqslant 1, \end{cases}$$

where $\lambda' := [6 + 4\pi(1-\pi)]/[15 + 4\pi(1-\pi)]$, $\lambda'' := [12 + 8\pi(1-\pi)]/[21 + 26\pi(1-\pi)]$ and $\lambda''' := 4/7$.

□

Proof of Lemma 1 (Complete information). We begin with player 1, and then consider an interior player i. We exclude player N: knowledge of the state is irrelevant for her so that her behaviour continues to be characterised by Lemma 1.

Player 1 First, note that cases $(ii) - (v)$ are not applicable to player 1 as $M_1 = 0$. Given the symmetric environment, we set $\theta_1 = 1$ without loss of generality. Suppose first $\theta_1 = \omega$. Because $\mathbb{E}[\bar{s}_{-1}|M_1, 1; \omega = 1] = \mathbb{E}[\bar{s}_{-1}|\omega = 1] = \pi$, a player 1 with $\lambda_1 \geqslant \lambda_1(M_1, 1; \omega = 1) = (1-2\pi)/(2-2\pi)$ chooses to reveal her type. Since $\pi > 1/2$ implies $\lambda_1(M_1, 1; \omega = 1) < 0$, player 1 always chooses $s_1^*(\theta_1) = \theta_1$, proving case (i).

Suppose now $\theta_1 \neq \omega$. Because $\mathbb{E}[\bar{s}_{-1}|M_1, 1; \omega = 0] = \mathbb{E}[\bar{s}_{-1}|\omega = 0] = 1 - \pi$, a player 1 with $\lambda_1 \geqslant \lambda_1(M_1, 1; \omega = 0) = (2\pi - 1)/(2\pi)$ chooses to reveal her type. Since $\pi > 1/2$ implies $\lambda_1(M_1, 1; \omega = 0) > 0$, a player 1 with $\lambda_1 \geqslant \lambda_1(M_1, \theta_1 \neq \omega)$ chooses $s_1^*(\theta_1) = \theta_1$ whilst a player 1 with $\lambda_1 < \lambda_1(M_1, \theta_1 \neq \omega)$ chooses $s_1^*(\theta_1) = 1 - \theta_1$, proving case (vi).

Player i We now consider player $i \in \{2, \ldots, N-1\}$. As above, we set $\theta_i = 1$ without loss of generality. Suppose that $\theta_i = \omega$. Given that $\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 1] = M_i/(N-1) + \pi(N-i)/(N-1)$,

$$\lambda_i(M_i, 1; \omega = 1) = \frac{1 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 1]}{2 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 1]} = \frac{N - 1 - 2[M_i + \pi(N-i)]}{2(N-1) - 2[M_i + \pi(N-i)]}. \quad \text{(B.25)}$$

The denominator of (B.25) is always positive. The numerator is positive for $\pi < \underline{\pi}$, where $\underline{\pi} := [N - 1 - 2M_i]/[2(N-i)]$. Note that $\underline{\pi} \leqslant 1/2$ rearranges to $M_i \geqslant (i-1)/2$. Therefore, case (i) implies $\lambda_i(M_i, \theta_i = \omega) < 0$ so that player i always chooses $s_i^*(\theta_i) = \theta_i$. Note that $\underline{\pi} \geqslant 1$ arranges to $M_i \leqslant (2i - N - 1)/2$. Therefore, case (iii) implies $\lambda_i(M_i, \theta_i = \omega) > 0$ so that a player i with $\lambda_i \geqslant \lambda_i(M_i, \theta_i = \omega)$ chooses $s_i^*(\theta_i) = \theta_i$ whilst a player i with $\lambda_i < \lambda_i(M_i, \theta_i = \omega)$ chooses $s_i^*(\theta_i) = 1 - \theta_i$. Finally, $1/2 < \underline{\pi} < 1$ rearranges to the conditions in case (ii) so that a player i with $\lambda_i < \lambda_i(M_i, \theta_i = \omega)$ chooses $s_i^*(\theta_i) = 1 - \theta_i$ in an uninformative environment $\pi < \underline{\pi}$. She chooses $s_i^*(\theta_i) = \theta_i$ otherwise.

Type realisation		**Outcome**				**Welfare W (ex-post)**			
$(\theta_1, \theta_2, \theta_3, \theta_4)$	$\Pr(\theta_1, \theta_2, \theta_3, \theta_4)$	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
$(0,0,0,0)$	$[(1-\pi)^4+\pi^4]/2$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	0	0	0	0
$(0,0,0,1)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,0,1)$	$-\lambda$	$-\lambda$	$-\lambda$	$-(1-\lambda)4/3$
$(0,0,1,0)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,1,0)$	$(0,0,1,0)$	$-\lambda$	$-\lambda$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(0,0,1,1)$	$(1-\pi)^2\pi^2$	$(0,0,0,0)$	$(0,0,0,0)$	$(0,0,1,1)$	$(0,0,1,1)$	-2λ	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(0,1,0,0)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(0,0,0,0)$	$(0,1,0,0)$	$(0,1,0,0)$	$(0,1,0,0)$	$-\lambda$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(0,1,0,1)$	$(1-\pi)^2\pi^2$	$(0,0,0,0)$	$(0,1,0,1)$	$(0,1,0,1)$	$(0,1,0,1)$	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(0,1,1,0)$	$(1-\pi)^2\pi^2$	$(0,0,0,0)$	$(0,1,1,0)$	$(0,1,1,0)$	$(0,1,1,0)$	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(0,1,1,1)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(0,0,0,0)$	$(0,1,1,1)$	$(0,1,1,1)$	$(0,1,1,1)$	-3λ	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(1,0,0,0)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(1,1,1,1)$	$(1,0,0,0)$	$(1,0,0,0)$	$(1,0,0,0)$	-3λ	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(1,0,0,1)$	$(1-\pi)^2\pi^2$	$(1,1,1,1)$	$(1,0,0,1)$	$(1,0,0,1)$	$(1,0,0,1)$	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(1,0,1,0)$	$(1-\pi)^2\pi^2$	$(1,1,1,1)$	$(1,0,1,0)$	$(1,0,1,0)$	$(1,0,1,0)$	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(1,0,1,1)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(1,1,1,1)$	$(1,0,1,1)$	$(1,0,1,1)$	$(1,0,1,1)$	$-\lambda$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(1,1,0,0)$	$(1-\pi)^2\pi^2$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,0,0)$	$(1,1,0,0)$	-2λ	-2λ	$-(1-\lambda)16/9$	$-(1-\lambda)16/9$
$(1,1,0,1)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,0,1)$	$(1,1,0,1)$	$-\lambda$	$-\lambda$	$-(1-\lambda)4/3$	$-(1-\lambda)4/3$
$(1,1,1,0)$	$[(1-\pi)^3\pi+\pi^3(1-\pi)]/2$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,1,0)$	$-\lambda$	$-\lambda$	$-\lambda$	$-(1-\lambda)4/3$
$(1,1,1,1)$	$[(1-\pi)^4+\pi^4]/2$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,1,1)$	$(1,1,1,1)$	0	0	0	0

Table B.3: Welfare calculation for mappings $(i)-(iv)$.

Suppose now that $\theta_i \neq \omega$. Given that $\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 0] = M_i/(N-1) + (1-\pi)(N-i)/(N-1)$,

$$\lambda_i(M_i, 1; \omega = 0) = \frac{1 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 0]}{2 - 2\mathbb{E}[\bar{s}_{-i}|M_i, 1; \omega = 0]} = \frac{N - 1 - 2[M_i + (1-\pi)(N-i)]}{2(N-1) - 2[M_i + (1-\pi)(N-i)]}. \quad \text{(B.26)}$$

The denominator of (B.26) is always positive. The numerator is positive for $\pi > \bar{\pi}$, where $\bar{\pi} := [N + 1 - 2(i - M_i)]/[2(N-i)]$. In this case, $\bar{\pi} \geqslant 1$ rearranges to $M_i \geqslant (N-1)/2$, while $\bar{\pi} \leqslant 1/2$ rearranges to $M_i \leqslant (i-1)/2$. The behaviour of player i then follows from above reasoning. □

Proof of Lemma 2 (Complete information). Let player i have one successor. Then, we can write her utility from truth telling as $\mathbb{E}[U(\theta_i, \theta_i)] = -pu_1(1-\lambda_i) - (1-p)u_2(1-\lambda_i)$, where

$$u_1 := \left(\theta_i - \frac{s_1 + \ldots + s_{i-1} + \theta_i}{N-1}\right)^2 < \left(\theta_i - \frac{s_1 + \ldots + s_{i-1} + 1 - \theta_i}{N-1}\right)^2 =: u_2,$$

and her utility from falsifying as $\mathbb{E}[U(1-\theta_i, \theta_i)] = -qz_1(1-\lambda_i) - (1-q)z_2(1-\lambda_i) - \lambda_i$, where

$$z_1 := \left(1 - \theta_i - \frac{s_1 + \ldots + s_{i-1} + \theta_i}{N-1}\right)^2 > \left(1 - \theta_i - \frac{s_1 + \ldots + s_{i-1} + 1 - \theta_i}{N-1}\right)^2 =: z_2.$$

Note that $\mathbb{E}[U(\theta_i, \theta_i)] > \mathbb{E}[U(1-\theta_i, \theta_i)]$ is equivalent to $\lambda_i < (u-z)/(1+u-z)$, where $z := qz_1 + (1-q)z_2$ and $u := pu_1 + (1-p)u_2$.

A credulous or sceptical player i posits her successor to be truthful independently of her action so that under complete information $p = q = \pi$ when $\theta_i = \omega$, and $p = q = 1 - \pi$ when $\theta_i \neq \omega$. A rational player i instead takes into account that her successor agrees with the majority whenever the history is sufficiently uniform: $p = 1$ if $(M_i + 1)/i > x$, and $q = 1$ if $M_i/i > x$; conversely, $p = 0$ if $(M_i + 1)/i < 1 - x$, and $q = 0$ if $M_i/i < 1 - x$, where $x \in [1/2, 1]$.

Case 1 Consider a sufficiently uniform history so that player i's action does not affect her successor's behaviour. If $\min\{M_i/i, (M_i+1)/i\} = M_i/i > x$, $u_R = u_1 < pu_1 + (1-p)u_2 = u_{CS}$ while $z_R = z_1 > qz_1 + (1-q)z_2 = z_{CS}$. Therefore, $u_R - z_R < u_{CS} - z_{CS}$ so that

$$\lambda_i^R := \frac{u_R - z_R}{1 + u_R - z_R} < \frac{u_{CS} - z_{CS}}{1 + u_{CS} - z_{CS}} =: \lambda_i^{CS}.$$

If $\max\{M_i/i, (M_i+1)/i\} = (M_i+1)/i < 1 - x$, $u_R > u_{CS}$ while $z_R < z_{CS}$. As $u_R - z_R > u_{CS} - z_{CS}$, $\lambda_i^R > \lambda_i^{CS}$.

Case 2 Consider a sufficiently diverse history so that player i's action does not affect her successor's behaviour. If $(M_i+1)/i \leqslant x$ and $M_i/i \geqslant 1 - x$, $u_R = u_{CS}$ and $z_R = z_{CS}$ so that $\lambda_i^R = \lambda_i^{CS}$.

Case 3 Consider a *non-tied* history so that player i's action does affect her successor's behaviour. If $M_i/i = x$ and $(M_i+1)/i > x$, $z_R = z_{CS}$ and $u_R = u_1 < pu_1 + (1-p)u_2 = u_{CS}$. Therefore, $u_R - z_R < u_{CS} - z_{CS}$ so that $\lambda_i^R < \lambda_i^{CS}$. If $(M_i+1)/i = 1-x$ and $M_i/i < 1-x$, $u_R = u_{CS}$ and $z_R = z_2 < qz_1 + (1-q)z_2 = z_{CS}$. Therefore, $u_R - z_R > u_{CS} - z_{CS}$ so that $\lambda_i^R > \lambda_i^{CS}$.

Case 1 and 3 together imply $\lambda_i^R < \lambda_i^{CS}$ for $M_i \geqslant ix > (i-1)/2$, and $\lambda_i^R > \lambda_i^{CS}$ for $M_i \leqslant i(1-x) - 1 < (i-1)/2$.

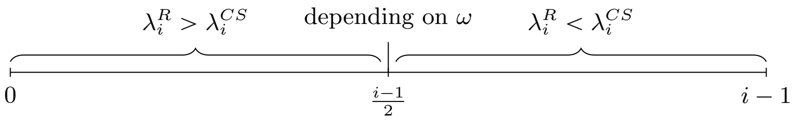
Case 4 Consider a tied history, $M_i = (i-1)/2$, for odd player i. If her successor is not too conformist, $x \geqslant (i+1)/(2i)$, **Case 2** applies with $\lambda_i^R = \lambda_i^{CS}$. Then, Figure B.5b is the partition of M_i.

If $x < (i+1)/(2i)$,[2] the successor always agrees with player i's action so that $p = 1$ and $q = 0$. In this case, $u_R = u_1$ and $z_R = z_2$. As $u_{CS} > u_R$ and $z_{CS} > z_R$, $u_R - z_R \lesseqgtr u_{CS} - z_{CS}$. We can make progress using $M_i = (i-1)/2$ and $i = N-1$. Without loss of generality, set $\theta_i = 1$ and substitute:

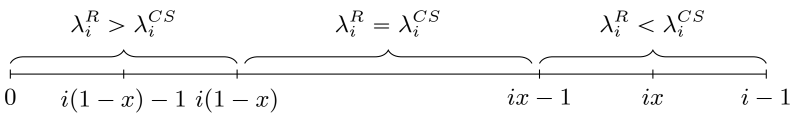
$$\begin{aligned}
u_1 &= \left(1 - \frac{M_i+1}{N-1}\right)^2 = \left(1 - \frac{i+1}{2(N-1)}\right)^2 = \left(\frac{N-2}{2(N-1)}\right)^2, \\
z_2 &= \left(0 - \frac{M_i+0}{N-1}\right)^2 = \left(\frac{i-1}{2(N-1)}\right)^2 = \left(\frac{N-2}{2(N-1)}\right)^2, \\
u_2 &= \left(1 - \frac{M_i+0}{N-1}\right)^2 = \left(1 - \frac{i-1}{2(N-1)}\right)^2 = \left(\frac{N}{2(N-1)}\right)^2, \text{ and} \\
z_1 &= \left(0 - \frac{M_i+1}{N-1}\right)^2 = \left(\frac{i+1}{2(N-1)}\right)^2 = \left(\frac{N}{2(N-1)}\right)^2.
\end{aligned}$$

Therefore, $u_{CS} = pu_1 + (1-p)u_2$ and $z_{CS} = qz_1 + (1-q)z_2 = pu_2 + (1-p)u_1$ so that $u_{CS} - z_{CS} = (1-2p)(u_2 - u_1)$. As $u_2 > u_1$, the first factor determines the sign. If $\theta_i = \omega$, $p = \pi > 1/2$ so that $u_{CS} - z_{CS} < 0 = u_R - z_R$. If $\theta_i \neq \omega$, $p = 1 - \pi < 1/2$ so that $u_{CS} - z_{CS} > 0 = u_R - z_R$. This implies that $\lambda_i^{CS} < 0 = \lambda_i^R$ if $\theta_i = \omega$, and $\lambda_i^{CS} > 0 = \lambda_i^R$ if $\theta_i \neq \omega$. If the history ties and the successor is highly conformist, the state determines whether a rational player i's cut-off is higher or lower. Then, Figure B.5b is the partition of M_i.

[2] When N is large, $x = 1/2$ is essentially the only admissible value for $i = N-1$.



(a) Highly conformist successor.



(b) Less conformist successor.

Figure B.5: Partition of $M_i \in \{0, \ldots, i-1\}$.

□

Proof of Theorem 1 (Complete information). We prove claim (i) and (ii) separately.

(i) The cut-offs follow from the proof of Lemma 1 (Complete information): $\lambda_i(0, \theta_i \neq \omega) = \frac{N-1-2(1-\pi)(N-i)}{2(N-1)-2(1-\pi)(N-i)} > 0$ and $\lambda_i(i-1, \theta_i = \omega) = \frac{2(N-i)(1-\pi)-(N-1)}{2(N-i)(1-\pi)} < 0$. Given that $\lambda_i(0, \theta_i \neq \omega)$ is increasing in i, $\lambda < (2\pi - 1)/(2\pi) = \lambda_1(0, \theta_1 \neq \omega)$ is sufficient for player 1 and all later players with $\theta_i \neq \omega$ to falsify.

(ii) For non-existence of the mapping under homogeneous conformity motives, we require the interval $\lambda_1(0, \theta_1 \neq \omega) \leqslant \lambda < \lambda_2(0, \theta_2 = \omega)$ to be empty. As $\partial\lambda_1(0, \theta_1 \neq \omega)/\partial\pi > 0$ and $\partial\lambda_2(0, \theta_2 = \omega)/\partial\pi < 0$ on $(1/2, 1)$, the two cut-offs cross at most once. As $\lim_{\pi \downarrow 1/2} \lambda_1(0, \theta_1 \neq \omega) = 0$ and $\lim_{\pi \uparrow 1} \lambda_1(0, \theta_1 \neq \omega) = 1/2$, whilst $\lim_{\pi \downarrow 1/2} \lambda_2(0, \theta_2 = \omega) = 1/N > 0$ and $\lim_{\pi \uparrow 1} \lambda_2(0, \theta_2 = \omega) = -\infty$, the cut-offs cross exactly once on $(1/2, 1)$. Hence, the interval is empty when $\lambda_1(0, \theta_1 \neq \omega) \geqslant \lambda_2(0, \theta_2 = \omega)$ which rearranges to $\pi \geqslant (N-1)/(2N-3)$. □

Bibliography

Abeles, Hal, "Are Musical Instrument Gender Associations Changing?," *Journal of Research in Music Education*, 2009, *57* (2), 127–139.

Adler, Moshe, "Stardom and Talent," *The American Economic Review*, 1985, *75* (1), 208–212.

Ahler, Douglas J., "Self-Fulfilling Misperceptions of Public Polarization," *The Journal of Politics*, 2014, *76* (3), 607–620.

Akerlof, George A., "The Market for"Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Alaref, Jumana, Samira Nikaein Towfighian, Gustavo Nicolas Paez, and Mohammed Audah, *The Impact of Employer Discrimination on Female Labor Market Outcomes: Experimental Evidence from Tunisia, Policy Research Working Paper 9361*, Washington, DC: World Bank, 2020.

Allmendinger, Jutta and J. Richard Hackman, "The More, the Better? A Four-Nation Study of the Inclusion of Women in Symphony Orchestras," *Social Forces*, 1995, *74* (2), 423–460.

Antecol, Heather and Deborah A. Cobb-Clark, "Identity and Racial Harassment," *Journal of Economic Behavior & Organization*, 2008, *66* (3), 529–557.

Antle, John M., "The New Economics of Agriculture," *American Journal of Agricultural Economics*, 1999, *81* (5), 993–1010.

Anwar, Shamena, "Testing for Discrimination: Evidence from the Game Show Street Smarts," *Journal of Economic Behavior & Organization*, 2012, *81* (1), 268–285.

Argo, Jennifer J., Darren W. Dahl, and Andrea C. Morales, "Positive Consumer Contagion: Responses to Attractive Others in a Retail Context," *Journal of Marketing Research*, 2008, *45* (6), 690–701.

Arnot, Chris, Peter C. Boxall, and Sean B. Cash, "Do Ethical Consumers care about Price? A Revealed Preference Analysis of Fair Trade Coffee Purchases," *Canadian Journal of Agricultural Economics*, 2006, *54* (4), 555–565.

Arrow, Kenneth, "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, pp. 3–33.

Arrow, Kenneth J., "Information and the Organization of Industry," *Rivista Internazionale di Scienze Sociali*, 2017, *125* (2), 125–130.

Athey, Susan, Christopher Avery, and Peter Zemsky, "Mentoring and Diversity," *American Economic Review*, 2000, *90* (4), 765–786.

Aylward, David and Gabrielle Carey, "High-value Niche Production: What Australian Wineries might learn from a Bordeaux First Growth," *International Journal of Technology, Policy and Management*, 2009, *9* (4), 342–357.

Bagues, Manuel and Maria J. Perez-Villadoniga, "Do Recruiters prefer Applicants with similar Skills? Evidence from a Randomized Natural Experiment," *Journal of Economic Behavior & Organization*, 2012, *82* (1), 12–20.

Bain, Joe S., "A Note on Pricing in Monopoly and Oligopoly," *The American Economic Review*, 1949, *39* (2), 448–464.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Banks, Jeffrey S. and Joel Sobel, "Equilibrium Selection in Signaling Games," *Econometrica*, 1987, *55* (3), 647–661.

Baye, Michael R., Keith J. Crocker, and Jiandong Ju, "Divisionalization, Franchising, and Divestiture Incentives in Oligopoly," *The American Economic Review*, 1996, *86* (1), 223–236.

Beaty, Roger, Paul Seli, and Daniel Schacter, "Thinking about the past and future in daily life: an experience sampling study of individual differences in mental time travel," *Psychological Research*, 2019, *83* (4), 805–816.

Becker, Gary S., *The Economics of Discrimination*, 2 ed., Chicago: University of Chicago Press, 1971.

Benabou, Roland, "Groupthink: Collective Delusions in Organizations and Markets," *The Review of Economic Studies*, 2013, *80* (2), 429–462.

_ **and Jean Tirole**, "Laws and Norms," *NBER Working Paper (No. 17579)*, 2011.

Bernheim, B. Douglas, "A Theory of Conformity," *Journal of Political Economy*, 1994, *102* (5), 841–877.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 2004, *94* (4), 991–1013.

_ **, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney**, "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," *The Review of Economic Studies*, 2019, *86* (1), 191–239.

Bicchieri, Christina, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge, UK: Cambridge University Press, 2005.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bjørner, Thomas Bue, Lars Gårn Hansen, and Clifford S. Russell, "Environmental Labeling and Consumers' Choice - An Empirical Analysis of the Effect of the Nordic Swan," *Journal of Environmental Economics and Management*, 2004, *47* (3), 411–434.

Bohren, J. Aislinn, "Informational Herding with Model Misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Brewers Association, "Stats and Data: Craft Brewer Definition," https://www.brewersassociation.org/statistics-and-data/craft-brewer-definition/ n.d. Accessed: 14 January 2021.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

Busic-Sontic, Ante, Natalia V. Czap, and Franz Fuerst, "The Role of Personality Traits in Green Decision-Making," *Journal of Economic Psychology*, 2017, *62*, 313–328.

Cain Miller, Claire, "Is Blind Hiring the Best Hiring?," *The New York Times Magazine*, 25 February 2016. https://www.nytimes.com/2016/02/28/magazine/is-blind-hiring-the-best-hiring.html. Accessed: 24 February 2019.

Card, David, Jochen Kluve, and Andrea Weber, "Active Labour Market Policy Evaluations: A Meta-Analysis*," *The Economic Journal*, 2010, *120* (548), F452–F477.

_ , _ , and _ , "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 2017, *16* (3), 894–931.

Carfagna, Lindsey B., Emilie A. Dubois, Connor Fitzmaurice, Monique Y. Ouimette, Juliet B. Schor, Margaret Willis, and Thomas Laidley, "An Emerging Eco-Habitus: The Reconfiguration of High Cultural Capital Practices among Ethical Consumers," *Journal of Consumer Culture*, 2014, *14* (2), 158–178.

Carlsson, Fredrik, Jorge H. García, and Åsa Löfgren, "Conformity and the Demand for Environmental Goods," *Environmental Resource Economics*, 2010, *47*, 407–421.

Carney, Dan, "5 Things You probably didn't know about Aston Martin," *Maxim*, 13 July 2016. https://www.maxim.com/rides/5-things-to-know-aston-martin-2016-7. Accessed: 22 May 2020.

Carroll, Glenn R. and Dennis Ray Wheaton, "The Organizational Construction of Authenticity: An Examination of Contemporary Food and Dining in the U.S.," *Research in Organizational Behavior*, 2009, *29*, 255–282.

Castiglione, Lawrence V., "Performance, Practice, and Policy Considerations," *Design For Arts in Education*, 1985, *1* (87), 31–37.

Centola, Damon, Robb Willer, and Michael Macy, "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms," *American Journal of Sociology*, 2005, *110* (4), 1009–1040.

Chaiken, Shelly and Yaacov Trope, *Dual-process Theories in Social Psychology*, New York; London: Guildford, 1999.

Chamberlin, Edward H., *The Theory of Monopolistic Competition: A Re-Orientation of the Theory of Value*, Cambridge, MA: Harvard University Press, 1933.

Champsaur, Paul and Jean-Charles Rochet, "Multiproduct Duopolists," *Econometrica*, 1989, *57* (3), 533–557.

Cho, In-Koo and David M. Kreps, "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, 1987, *102* (2), 179–221.

Church, E. Mitchell and Richelle L. Oakley, "Etsy and the Long-tail: How Microenterprises use Hyper-differentiation in Online Handicraft Marketplaces," *Electronic Commerce Research*, 2018, *18* (4), 883–898.

Cialdini, Robert B., *Influence: The Psychology of Persuasion*, New York: William Morrow and Company, 1984.

_ , Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter, "Managing Social Norms for Persuasive Impact," *Social Influence*, 2006, *1* (1), 3–15.

Coate, Stephen and Glenn C. Loury, "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *The American Economic Review*, 1993, *83* (5), 1220–1240.

Conor, Bridget, Rosalind Gill, and Stephanie Taylor, "Gender and Creative Labour," *The Sociological Review*, 2015, *63* (S1), 1–22.

Cornell, Bradford and Ivo Welch, "Culture, Information, and Screening Discrimination," *Journal of Political Economy*, 1996, *104* (3), 542–71.

Cornell Law School LII, "U.S. Supreme Court - GRATZ v. Bollinger," https://www.law.cornell.edu/supremecourt/text/02-516 n.d. Accessed: 4 June 2022.

Daskin, Alan J., "Two Propositions on the Application of Point Elasticities to Finite Price Changes," *The Journal of Economic Education*, 1992, *23* (1), 17–21.

d'Aspremont, Claude, Jean Jaskold Gabszewicz, and Jaques-François Thisse, "On Hotelling's "Stability in Competition"," *Econometrica*, 1979, *47* (5), 1145–1150.

Davidson, Adam, "Are We in Danger of a Beer Monopoly?," *The New York Times Magazine*, 26 February 2013. https://www.nytimes.com/2013/03/03/magazine/beer-mergers.html. Accessed: 15 January 2021.

De Pelsmacker, Patrick and Wim Janssens, "A Model for Fair Trade Buying Behaviour: The Role of Perceived Quantity and Quality of Information and of Product-specific Attitudes," *Journal of Business Ethics*, 2007, *75* (4), 361–380.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, June 2009, *47* (2), 315–72.

Deneckere, Raymond J. and R. Preston McAfee, "Damaged Goods," *Journal of Economics & Management Strategy*, 1996, *5* (2), 149–174.

Dickson, Marsha A. and Mary A. Littrell, "Consumers of Ethnic Apparel from Alternative Trading Organizations: A Multifaceted Market," *Clothing and Textiles Research Journal*, 1998, *16* (1), 1–10.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper, 1957.

Dragusanu, Raluca, Daniele Giovannucci, and Nathan Nunn, "The Economics of Fair Trade," *Journal of Economic Perspectives*, 2014, *28* (3), 217–36.

Duff, William, *An Essay on Original Genius; and its Various Modes of Exertion in Philosophy and the Fine Arts, particularly in Poetry*, London: E. and C. Dilly, 1767.

Eaton, B. Curtis and Richard G. Lipsey, "Product Differentiation," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, Vol. 1, Amsterdam: Elsevier, 1989, chapter 12, pp. 723–768.

Economides, Nicholas, "Quality Variations and Maximal Variety Differentiation," *Regional Science and Urban Economics*, 1989, *19* (1), 21–29.

_ , "Quality Variations in the Circular Model of Variety-differentiated Products," *Regional Science and Urban Economics*, 1993, *23* (2), 235–257.

Elliott, Charles A., "Race and Gender as Factors in Judgments of Musical Performance," *Bulletin of the Council for Research in Music Education*, 1995, *127* (Winter), 50–56.

Ericsson, K. Anders, "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance," in K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, eds., *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge: Cambridge University Press, 2006, chapter 38, pp. 683–703.

_ **, Ralf Th. Krampe, and Clemens Tesch-Romer**, "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review*, 1993, *3* (100), 363–406.

Esgate, Anthony and David Groome, *An Introduction to Applied Cognitive Psychology*, Hove, UK; New York: Psychology Press, 2005.

Etsy, Inc., "2018 Annual Report," https://s22.q4cdn.com/941741262/files/doc_financials/annual/2018-Annual-Report-(1).pdf 28 February 2019. Accessed: 23 January 2020.

—, "Handmade Policy - Our House Rules," https://www.etsy.com/uk/legal/handmade 9 April 2019. Accessed: 9 January 2021.

—, "First Quarter 2015 - Third Quarter 2020 Financial Results," https://investors.etsy.com/press-releases n.d. Accessed: 5 January 2021.

Eyster, Erik and Matthew Rabin, "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2010, *2* (4), 221–43.

Frey, Bruno S. and Stephan Meier, "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment," *The American Economic Review*, 2004, *94* (5), 1717–1722.

Friedrichsen, Jana and Dirk Engelmann, "Who cares about Social Image?," *European Economic Review*, 2018, *110*, 61–77.

Fryer, Roland and Matthew Jackson, "A Categorical Model of Cognition and Biased Decision Making," *The B.E. Journal of Theoretical Economics*, 2008, *8* (1), 1–44.

Fuchs, Christoph, Martin Schreier, and Stijn M.J. van Osselaer, "The Handmade Effect: What's Love Got to Do with It?," *Journal of Marketing*, 2015, *79* (2), 98–110.

Fudenberg, Drew and Jean Tirole, *Game Theory*, Cambridge, MA: MIT Press, 1991.

Galarraga, Ibon and Anil Markandya, "Economic Techniques to estimate the Demand for Sustainable Products: A Case Study for Fair Trade and Organic Coffee in the United Kingdom," *Economia Agraria y Recursos Naturales, Spanish Association of Agricultural Economists*, 2004, *4* (7), 1–26.

Galenson, David W., *Old Masters and Young Geniuses: The Two Life Cycles of Artistic Creativity*, Princeton University Press, 2006.

—, "Old masters and young geniuses: The two life cycles of human creativity," *Journal of Applied Economics*, 2009, *12* (1), 1–9.

Gallwey, W. Timothy, *The Inner Game of Tennis*, New York: Random House, 1974.

Geiger, Nathaniel and Janet K. Swim, "Climate of Silence: Pluralistic Ignorance as a Barrier to Climate Change Discussion," *Journal of Environmental Psychology*, 2016, *47*, 79–90.

Gigerenzer, Gerd, Peter Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart*, Oxford: Oxford University Press, 1999.

Gilbert, Richard J., "The Role of Potential Competition in Industrial Organization," *Journal of Economic Perspectives*, 1989, *3* (3), 107–127.

Gioia, Dennis A. and Henry P. Sims Jr., "Self-serving Bias and Actor-Observer Differences In Organizations: An Empirical Analysis," *Journal of Applied Social Psychology*, 1985, *15* (6), 547–563.

Gladwell, Malcolm, *Blink: The Power of Thinking without Thinking*, New York: Little, Brown & Company, 2005.

Glazebrook, Kate, "What is Blind Hiring?," *Applied*, 22 July 2019. https://www.beapplied.com/post/what-is-blind-hiring. Accessed: 18 November 2019.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *The American Economic Review*, 2000, *90* (4), 715–741.

Granville, Brigitte and Steven Telford, "Empirical Evidence from South Africa," in Brigitte Granville and Janet Dine, eds., *The Processes and Practices of Fair Trade: Trust, Ethics and Governance.*, New York, NY: Routledge, 2013, chapter 10, pp. 286–340.

Griffiths, Noola K., "The Effects of Concert Dress and Physical Appearance on Perceptions of Female Solo Performers," *Musicae Scientiae*, 2008, *12* (2), 273–290.

Grossman, Sanford J., "An Introduction to the Theory of Rational Expectations Under Asymmetric Information," *Review of Economic Studies*, 1981, *48* (4), 541–559.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel, "Women can't jump?—An Experiment on Competitive Attitudes and Stereotype Threat," *Journal of Economic Behavior & Organization*, 2010, *75* (3), 395–401.

Harsanyi, John C., "Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model," *Management Science*, 1967, *14* (3), 159–182.

— , "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part II. Bayesian Equilibrium Points," *Management Science*, 1968, *14* (5), 320–334.

— , "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part III. The basic probability distribution of the game," *Management Science*, 1968, *14* (5), 486–502.

Hart, Philip, *Orpheus in the New World: The Symphony Orchestras as an American Cultural Institution— its Past, Present, and Future*, New York: W.W. Norton & Company, 1973.

Hays, Kate F. and Charles H. Brown, *You're on! Consulting for Peak Performance*, Washington, DC: American Psychological Association, 2006.

Hellmanzik, Christiane, *Cultural Economics*, Newcastle upon Tyne: Agenda Publishing, 2020.

Hertwig, Ralph and Till Grüne-Yanoff, "Nudging and Boosting: Steering or Empowering Good Decisions," *Perspectives on Psychological Science*, 2017, *12* (6), 973–986.

Hewstone, Miles, Mark Rubin, and Hazel Willis, "Intergroup Bias," *Annual Review of Psychology*, 2002, *53* (1), 575–604.

Hill, Andrew, "Why Groupthink Never Went Away," *Financial Times*, 6 May 2018. https://www.ft.com/content/297ffe7c-4ee4-11e8-9471-a083af05aea7. Accessed: 13 February 2020.

Holland, Bernard, "The Fair, New World of Orchestra Auditions," *The New York Times*, 11 January 1981.

Holmström, Bengt, "Moral Hazard and Observability," *The Bell Journal of Economics*, 1979, *10* (1), 74–91.

ILO, *Ergonomic Checkpoints: Practical and Easy-to-implement Solutions for Improving Safety, Health and Working Conditions*, 2 ed., Geneva: International Labour Organization, 2010.

Internicola, Dorene, "Hoop dreams: cashing in on hula craze," *Reuters*, 21 January 2009. https://www.reuters.com/article/us-fitness-hooping-idUSTRE50K5GF20090121. Accessed: 28 May 2022.

Jakob, Doreen, "Crafting your way out of the recession? New craft entrepreneurs and the global economic downturn," *Cambridge Journal of Regions, Economy and Society*, 2012, *6* (1), 127–140.

_ **and Nicola J. Thomas**, "Firing up craft capital: the renaissance of craft and craft policy in the United Kingdom," *International Journal of Cultural Policy*, 2017, *23* (4), 495–511.

Janis, Irving L., *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*, Boston: Houghton Mifflin, 1972.

_ , "Groupthink," in Emory A. Griffin, ed., *A First Look at Communication Theory*, McGrawHill, 1991, pp. 235–246.

Jensen, Michael C. and William H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 1976, *3*, 305–360.

Johnson, Justin P. and David P. Myatt, "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 2006, *96* (3), 756–784.

Johnston, José and Shyon Baumann, "Democracy versus Distinction: A Study of Omnivorousness in Gourmet Food Writing," *American Journal of Sociology*, 2007, *113* (1), 165–204.

Jones, Edward E. and Richard E. Nisbett, *The actor and the observer: Divergent perceptions of the causes of*, Morristown, N.J.: General Learning Press, 1971.

_ **and** _ , "The Actor and the Observer: Divergent Perceptions of the Causes of Behavior," in Edward E. Jones, David E. Kanouse, Harold H. Kelley, Richard E. Nisbett, Stuart Valins, and Bernard Weiner, eds., *Attribution: Perceiving the Causes of Behavior*, Lawrence Erlbaum Associates, Inc, 1987, p. 79–94.

Kahan, Dan M., "Ideology, motivated reasoning, and cognitive reflection," *Judgment and Decision Making*, 2013, *8* (4), 407–424.

Kahneman, Daniel, *Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice*, Stockholm: The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2002, 8 December 2002.

Kamakura, Wagner A. and Carl W. Schimmel, "Uncovering Audience Preferences for Concert Features from Single-Ticket Sales with a Factor-Analytic Random-Coefficients Model," *International Journal of Research in Marketing*, 2013, *30* (2), 129–142.

Kets, Willemien and Alvaro Sandroni, "A Theory of Strategic Uncertainty and Cultural Diversity," *The Review of Economic Studies*, 2021, *88* (1), 287–333.

Kreuzbauer, Robert, Dan King, and Shankha Basu, "The Mind in the Object-Psychological Valuation of Materialized Human Expression," *Journal of Experimental Psychology*, 2015, *144*, 764–787.

Krumbiegel, Katharina, Miet Maertens, and Meike Wollni, "The Role of Fairtrade Certification for Wages and Job Satisfaction of Plantation Workers," *World Development*, 2018, *102*, 195–212.

Kuran, Timur, "Chameleon Voters and Public Choice," *Public Choice*, 1987, *53* (1), 53–78.

_ , "Preference Falsification, Policy Continuity and Collective Conservatism," *The Economic Journal*, 1987, *97* (387), 642–665.

_ , *Private Truths, Public Lies: The Social Consequences of Preference Falsification.*, Cambridge, MA: Harvard University Press, 1995.

_ **and Cass R. Sunstein**, "Availability Cascades and Risk Regulation," *Stanford Law Review*, 1999, *51* (4), 683–768.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94* (3), 614–641.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

Lancaster, Kelvin J., "A New Approach to Consumer Theory," *Journal of Political Economy*, 1966, *74* (2), 132–157.

Lehman, David W., Kieran O'Connor, and Glenn R. Carroll, "Acting on Authenticity: Individual Interpretations and Behavioral Responses," *Review of General Psychology*, 2019, *23* (1), 19–31.

Leland, Hayne, "Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards," *Journal of Political Economy*, 1979, *87* (6), 1328–1346.

Levinson, Gary, "Preparing for an Orchestral Audition: the Basics," *The Strad*, 26 July 2017. https://www.thestrad.com/preparing-for-an-orchestral-audition-the-basics/6964.article. Accessed: 22 February 2019.

Liebl, Maureen and Tirthankar Roy, "Handmade in India: Preliminary Analysis of Crafts Producers and Crafts Production," *Economic and Political Weekly*, 2003, *38* (51/52), 5366–5376.

Loureiro, Maria L. and Justus Lotade, "Do Fair Trade and Eco-Labels in Coffee wake up the Consumer Conscience?," *Ecological Economics*, 2005, *53* (1), 129–138.

Makoff-Clark, Annie, "Is Blind Recruitment the Secret to the Perfect Hire?," *People Management*, 24 January 2019. https://www.peoplemanagement.co.uk/long-reads/articles/blind-recruitment-secret-perfect-hire. Accessed: 14 November 2019.

Manzoni, John, "Name-Blind Recruitment – A Commitment to Diversity," *Civil Service Blog*, 5 November 2015. https://civilservice.blog.gov.uk/2015/11/05/name-blind-recruitment-a-commitment-to-diversity/. Accessed: 18 November 2019.

March, James G., "Bounded Rationality, Ambiguity, and the Engineering of Choice," *The Bell Journal of Economics*, 1978, *9* (2), 587–608.

Markoff, John, "Skilled Work, Without the Worker," *The New York Times*, 18 August 2012. http://www.nytimes.com/2012/08/19/business/new-wave-of-adept-robots-is-changing-global-industry.html?pagewanted=all. Accessed: 1 December 2019.

Martignon, Laura, Konstantinos V. Katsikopoulos, and Jan K. Woike, "Categorization with Limited Resources: A Family of Simple Heuristics," in Gerd Gigerenzer, Ralph Hertwig, and Thorsten Pachur, eds., *Heuristics: The Foundations of Adaptive Behavior*, Oxford University Press, 2011, chapter 14.

_ , Oliver Vitouch, Masanori Takezawa, and Malcolm R. Forster, "Naive and yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees," in Gerd Gigerenzer, Ralph Hertwig, and Thorsten Pachur, eds., *Heuristics: The Foundations of Adaptive Behavior*, Oxford University Press, 2011, chapter 6.

McPherson, Gary E. and Emery Schubert, "Measuring Performance Enhancement in Music," in Aaron Williamon, ed., *Musical Excellence: Strategies and Techniques to Enhance Performance*, Oxford: Oxford University Press, 2004, chapter 4, pp. 61–82.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 2001, *27* (1), 415–444.

Melloni, Gaia, Andrea Patacconi, and Nick Vikander, "CEO Activism as Communication to Multiple Audiences," *Working Paper*, 2019.

Meyer, Margaret A. and John Vickers, "Performance Comparisons and Dynamic Incentives," *Journal of Political Economy*, 1997, *105* (3), 547–581.

Mildenberger, Matto and Dustin Tingley, "Beliefs about Climate Beliefs: The Importance of Second-Order Opinions for Climate Politics," *British Journal of Political Science*, 2019, *49* (4), 1279–1307.

Milgram, Stanley, "Behavioral Study of Obedience.," *The Journal of Abnormal and Social Psychology*, 1963, *67* (4), 371–378.

Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 1981, *12* (2), 380–391.

Miller, Arthur G., Barry Gillen, Charles Schenker, and Shirley Radlove, "The Prediction and Perception of Obedience to Authority," *Journal of Personality*, 1974, *42* (1), 23–42.

Miller, Dale T. and Cathy McFarland, "Pluralistic Ignorance: When Similarity is Interpreted as Dissimilarity," *Journal of Personality and Social psychology*, 1987, *53* (2), 298–305.

_ **and Deborah A. Prentice**, "Collective Errors and Errors about the Collective," *Personality and Social Psychology Bulletin*, 1994, *20* (5), 541–550.

Mohan, Sushil, *Fair Trade Without the Froth: A Dispassionate Economic Analysis of 'Fair Trade'*, London: The Institute of Economic Affairs, 2010.

Mohliver, Aharon, Donal Crilly, and Aseem Kaul, "In Dubious Battle: Issue Salience, Polarization, and Corporate Social Counterpositioning," *Working Paper*, 2019.

Moro, Andrea and Peter Norman, "Affirmative Action in a Competitive Economy," *Journal of Public Economics*, 2003, *87* (3), 567–594.

Morris, Stephen, "Political Correctness," *Journal of Political Economy*, 2001, *109* (2), 231–265.

Musicians' Union, "Employment Agreements: Negotiated Rates," https://www.musiciansunion.org.uk/Home/Advice/Playing-Live/Orchestral-Work/Pay-Conditions/Employment-Agreements n.d. Accessed: 23 February 2019.

Mussa, Michael and Sherwin Rosen, "Monopoly and Product Quality," *Journal of Economic Theory*, 1978, *18* (2), 301–317.

Myerson, Roger B., "Comments on "Games with Incomplete Information Played by 'Bayesian' Players, I–III Harsanyi's Games with Incomplete Information"," *Management Science*, 2004, *50* (12_supplement), 1818–1824.

National Endowment For The Arts, "Art-Goers in Their Communities: Patterns of Civic and Social Engagement," *NEA Research Note No.98*, October 2009. https://www.arts.gov/sites/default/files/98.pdf. Accessed: 09 March 2019.

Newman, Leonard S. and James S. Uleman, "Spontaneous Trait Inference," in J. S. Uleman and J. A. Bargh, eds., *Unintended thought*, New York, NY: Guilford Press, 1989, chapter 5, pp. 155–188.

Nicholson, Heather, "8 Ways Fairtrade Farmers Protect the Environment," *Fairtrade Foundation*, 5 June 2018. https://www.fairtrade.org.uk/Media-Centre/Blog/2018/June/8-ways-Fairtrade-protects-the-environment. Accessed: 30 January 2020.

Nisbett, Richard E., Craig Caputo, Patricia Legant, and Jeanne Marecek, "Behavior as Seen by the Actor and as Seen by the Observer," *Journal of Personality and Social Psychology*, 1973, *27* (2), 154–164.

Nostalgia Central, "Tupperware Parties," https://nostalgiacentral.com/pop-culture/fads/tupperware-parties/ n.d. Accessed: 28 May 2022.

Oberholzer-Gee, Felix, "Nonemployment Stigma as Rational Herding: A Field Experiment," *Journal of Economic Behavior & Organization*, 2008, *65* (1), 30–40.

Oreskes, Naomi and Erik M. Conway, *Merchants of Doubt: How a Handful of Scientists obscured the Truth on Issues from Tobacco Smoke to Global Warming*, New York; London: Bloomsbury Press, 2010.

Ostrom, Elinor, "Collective Action and the Evolution of Social Norms," *The Journal of Economic Perspectives*, 2000, *14* (3), 137–158.

Page, Lionel and Katie Page, "Last shall be First: A Field Study of Biases in Sequential Performance Evaluation on the Idol Series," *Journal of Economic Behavior & Organization*, 2010, *73* (2), 186–198.

Pepall, Lynne, "Strategic Product Choice and Niche Markets," *Journal of Economics & Management Strategy*, 1992, *1* (2), 397–417.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 1972, *62* (4), 659–661.

Poret, Sylvaine, "Mainstreaming Fair Trade: A Discussion through the Lipton Tea Case," in Patricia Crifo, ed., *Corporate Social Responsibility: From Compliance to Opportunity?*, Palaiseau: Ed. de l'Ecole Polytechnique, 2010, pp. 189–206.

Purkiss, Sharon L. Segrest, Pamela L. Perrewé, Treena L. Gillespie, Bronston T. Mayes, and Gerald R. Ferris, "Implicit Sources of Bias in Employment Interview Judgments and Decisions," *Organizational Behavior and Human Decision Processes*, 2006, *101* (2), 152–167.

Rasmussen, Anne S. and Dorthe Berntsen, "The reality of the past versus the ideality of the future: emotional valence and functional differences between past and future mental time travel," *Memory & Cognition*, 2013, *41* (2), 187–200.

Read, Stephen J., "Constructing Causal Scenarios," *Journal of Personality and Social Psychology*, 1987, *52* (2), 288–302.

Reinstein, David and Joon Song, "Efficient Consumer Altruism and Fair Trade Products," *Journal of Economics & Management Strategy*, 2012, *21* (1), 213–241.

Renard, Marie-Christine, "Fair Trade: Quality, Market and Conventions," *Journal of Rural Studies*, 2003, *19* (1), 87–96.

Riach, Peter A. and Judith Rich, "An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market," *The B.E. Journal of Economic Analysis & Policy*, 2006, *6* (2), 1–20.

Richards, Jay W., "The 'Everything Handmade' Trend will curb Job Losses," *Wall Street Journal*, 21 July 2018. https://www.wsj.com/articles/the-everything-handmade-trend-will-curb-job-losses-1532126362. Accessed: 29 January 2020.

Richardson, Martin and Frank Stähler, "Fair Trade," *Economic Record*, 2014, *90* (291), 447–461.

Rosen, Sherwin, "The Economics of Superstars," *The American Economic Review*, 1981, *71* (5), 845–858.

Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro, "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions," *American Journal of Political Science*, 2016, *60* (3), 783–802.

Sebald, Alexander and Nick Vikander, "Optimal Firm Behavior with Consumer Social Image Concerns and Asymmetric Information," *Journal of Economic Behavior & Organization*, 2019, *167*, 311–330.

Seltzer, George, *Music Matters: The Performer and the American Federation of Musicians*, Mettuchen, NJ: Scarecrow Press, 1989.

Sergeant, Desmond Charles and Evangelos Himonides, "Orchestrated Sex: The Representation of Male and Female Musicians in World-Class Symphony Orchestras," *Frontiers in Psychology*, 2019, *10*, 1–18.

Shaked, Avner and John Sutton, "Relaxing Price Competition Through Product Differentiation," *The Review of Economic Studies*, 1982, *49* (1), 3–13.

Simon, Herbert A., "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 1955, *69* (1), 99–118.

Siniscalchi, Marciano and Pietro Veronesi, *Self-image Bias and Lost Talent, Discussion Paper 15621)*, London: CEPR, 2020.

Sliwka, Dirk, "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *American Economic Review*, 2007, *97* (3), 999–1012.

Snower, Dennis J., "The Low-Skill, Bad-Job Trap," *CEPR Discussion Paper Series, Centre for Economic Policy Research*, 1994, *999*, 1–21.

Social Impact of the Arts Project, "Culture and Social Wellbeing in New York City: Highlights of a Two-Year Research Project," https://repository.upenn.edu/siap_culture_nyc/2 February 2017. Accessed: 09 March 2019.

Sowell, Thomas, *Affirmative Action around the World: An Empirical Study*, New Haven, CT: Yale University Press, 2004.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

Spiegelhalter, David, Mike Pearson, and Ian Short, "Visualizing Uncertainty About the Future," *Science*, 2011, *333* (6048), 1393–1400.

Stahl, Dale O. and Paul W. Wilson, "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 1995, *10* (1), 218–254.

Starr, Susan, "The Prejudice Against Women," *Music Journal*, 1 March 1974, *32* (3), 14–15 and 28.

Steinrücken, Torsten and Sebastian Jaenichen, "The Fair Trade Idea: Towards an Economics of Social Labels," *Journal of Consumer Policy*, 2007, *30* (3), 201–217.

Stiefenhofer, Pascal, "Conspicuous Ethical Consumption," *Theoretical Economics Letters*, 2019, *9* (1), 1–8.

Stigler, George J., "The Economics of Information," *Journal of Political Economy*, 1961, *69* (3), 213–225.

— , *The Theory of Price*, New York: Macmillan, 1966.

Stiglitz, Joseph E., "The Theory of "Screening," Education, and the Distribution of Income," *The American Economic Review*, 1975, *65* (3), 283–300.

Stronsick, Lisa M., Samantha E. Tuft, Sara Incera, and Conor T. McLennan, "Masculine Harps and Feminine Horns: Timbre and Pitch Level influence Gender Ratings of Musical Instruments," *Psychology of Music*, 2018, *46* (6), 896–912.

Sun, Heshan, "A Longitudinal Study of Herd Behavior in the Adoption and Continued Use of Technology," *MIS Quarterly*, 2013, *37* (4), 1013–1041.

Taylor, Curtis R. and Huseyin Yildirim, "Subjective Performance and the Value of Blind Evaluation," *The Review of Economic Studies*, 2011, *78* (2), 762–794.

Taylor, Shelley E. and Susan T. Fiske, "Point of View and Perceptions of Causality," *Journal of Personality and Social psychology*, 1975, *32* (3), 439–445.

— **and** — , "Salience, Attention, and Attribution: Top of the Head Phenomena," in Leonard Berkowitz, ed., *Advances in Experimental Social Psychology*, Vol. 11, Academic Press, 1978, pp. 249–288.

Thaler, Richard, "Toward a positive theory of consumer choice," *Journal of Economic Behavior & Organization*, 1980, *1* (1), 39–60.

Thaler, Richard H., *Misbehaving: The Making of Behavioral Economics*, New York: W. W. Norton & Company, 2015.

— **and Cass R. Sunstein**, *Nudge: Improving decisions about health, wealth, and happiness.*, Yale: Yale University Press, 2008.

The Council of the EU, "Voting System - Unanimity," https://www.consilium.europa.eu/en/council-eu/voting-system/unanimity/ 28 January 2020. Accessed: 12 July 2021.

The Economist, "All Ears," *The Economist*, 30 November 1996, *341* (7994), 89–90 (UK 135–136).

Thorngate, Warren, Robyn M. Dawes, and Margaret Foddy, *Judging Merit*, New York: Psychology Press, 2010.

Throsby, Charles D. and Glenn A. Withers, *The Economics of the Performing Arts*, London: Edward Arnold Publishers, 1979.

Towse, Ruth, "Market Value and Artists' Earnings," in Arjo Klamer, ed., *The Value of Culture: On the Relationship between Economics and Arts*, Amsterdam: Amsterdam University Press, 1996, chapter 6, pp. 96–107.

_ , "Human Capital and Artists' Labour Markets," in Victor Ginsburgh and David Throsby, eds., *Handbook of the Economics of the Arts and Culture*, Vol. 1, Amsterdam: North Holland Elsevier, 2006, chapter 24, pp. 865–894.

_ , *A Textbook of Cultural Economics*, Cambridge, UK: Cambridge University Press, 2010.

Tsay, Chia-Jung, "Sight over Sound in the Judgment of Music Performance," *PNAS*, 2013, *110* (36), 14580–14585.

_ , "The vision heuristic: Judging music ensembles by sight alone," *Organizational Behavior and Human Decision Processes*, 2014, *124* (1), 24–33.

_ **and Mahzarin R. Banaji**, "Naturals and Strivers: Preferences and Beliefs about Sources of Achievement," *Journal of Experimental Social Psychology*, 2011, *47* (2), 460–465.

Tulving, Endel, "Memory and consciousness.," *Canadian Psychology*, 1985, *26* (1), 1–12.

Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, 1974, *185* (4157), 1124–1131.

_ **and** _ , "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, *211* (4481), 453–458.

Universal Music Group, "Universal Music Group signs Global Recording, Artist Services Agreement for New TV Music Sensation, 'The Voice of ...'," https://www.universalmusic.com/universal-music-group-signs-global-recording-artist-services-agreement-for-new-tv-music-sensation-the-voice-of/ 19 January 2011. Accessed: 22 February 2019.

U.S. Commission on Civil Rights, "Statement on Affirmative Action," *Clearinghouse Publication No. 54*, 1977, pp. 1–12.

U.S. Department of Justice, "Justice Department Reaches Settlement with Anheuser-Busch InBev and Grupo Modelo in Beer Case," https://www.justice.gov/opa/pr/justice-department-reaches-settlement-anheuser-busch-inbev-and-grupo-modelo-beer-case 19 April 2013. Accessed: 15 January 2021.

Vernos, Isabelle, "Quotas are Questionable," *Nature*, 7 March 2013, *495*, 39.

Walker, Rob, "Handmade 2.0," *The New York Times Magazine*, 16 December 2007. https://www.nytimes.com/2007/12/16/magazine/16Crafts-t.html?_r=2&oref=slogin&ref=magazine&pagewanted=all. Accessed: 23 January 2020.

Wei, Jong-Shin, "On Teaching Price Elasticity of Demand and Change in Revenue due to Price Change – A Synthesis with and without Calculus," *International Journal of Business and Economics*, 2013, *12* (1), 1–14.