# Portraits of Power: Facial Appearances and the Tacit Domain of Political Selection in China*

Junyan Jiang† Songpo Yang‡

**Abstract**

As the epitome of modern, rational organizations, bureaucracies are often believed to select candidates based on rules and reason. We argue that intuitive—and even instinctive—assessments of candidates' external appearances sometimes underpin seemingly rational and calculated decisions. Using a novel, AI-based algorithm that learns and reproduces human assessments of facial appearances at scale, we examine how perceived facial traits influence the careers of over 4,000 mid- and senior-level Chinese officials. We find that officials who look more competent, trustworthy, and less aggressive enjoy significantly better promotion prospect and lower purge risk than their peers. Warmth-related traits (e.g., trustworthiness and non-aggressiveness) are especially valued at higher-level promotions and for male candidates. Additional analyses, including conjoint experiments with real officials, demonstrate that appearances' influence over selection preferences is comparable to performance or political connections. These findings challenge the prevailing meritocratic and relation-based theories of bureaucratic selection and highlight the role of impressions in the working of government institutions.

*Authors are listed in alphabetical order. Both contributed to the project equally.

†Assistant Professor, Department of Political Science, Columbia University. Email: jj3160@columbia.edu.

‡Ph.D candidate, Department of International Relations, Tsinghua University. Email: yangsp21@mails.tsinghua.edu.cn.

# Introduction

The selection and promotion of government leaders is one of the most fundamental tasks in all political systems (Besley 2005; Key 1956; Madison 1788 [2008]; Manion 2023; Plato 1968). Decisions about who holds political or administrative offices not only profoundly impact the direction of policies and the quality of governance, but also reflect the key priorities and values espoused by different regimes. A burgeoning body of scholarship has shown that in democracies, voters make electoral choices based not only on candidates' professional credentials or issue positions (Dal Bó and Finan 2018), but also on subjective assessments of style and personality—often derived from voters' perceptions of candidates' physical appearances (Antonakis and Dalgas 2009; Rule et al. 2010; Todorov 2017; Zebrowitz and Montepare 2005).[1] Researchers have found that looking attractive, competent, or simply resembling a stereotypical political leader provides candidates with a significant advantage at the ballot box (Banducci et al. 2008; Lawson et al. 2010; Lenz and Lawson 2011; Little et al. 2007; Mattes et al. 2010; Olivola and Todorov 2010a; Sigelman, Sigelman, and Fowler 1987; Todorov et al. 2005).

If voting is inherently susceptible to the influence of candidates' physical appeal due to the public-facing nature of elections and the informational and cognitive limitations of ordinary citizens, what about selection processes in more institutionalized, elite-dominated contexts? Besides elections, another common way to select officeholders is through top-down appointment by superiors within a government or party bureaucracy. Unlike popular elections, selection in a bureaucratic system is typically governed by elaborate rules and procedures and involves a small group of internal decision-makers who supposedly possess expert knowledge about candidates' work and qualifications (Downs 1967; Weber 1921 [1978]; Wilson 1989). Consequently, one might expect bureaucratic selection to minimize the influence of superficial traits such as appearances (Wong and Zeng 2017). In line with this view, some authors even go so far as to claim that bureaucratic selection represents a close approximation to the ideal of meritocracy, as it is conducted by seasoned insiders who are ostensibly free from the biases and limitations of lay voters (Bell 2016;

[1] For a recent review of studies on this topic, see Giacomin and Rule (2020).

Fukuyama 2014). However, a contrasting view suggests that elites and experts are just as likely to rely on impressions and instincts in their judgment as the general public (Gigerenzer 2007; Kahneman and Klein 2009), and the presence of considerable information asymmetry in bureaucratic environments often necessitates the use of heuristics and shortcuts in decision making (Barnard 1938; Simon 1947).

To better understand whether and how physical appearances influence selection in government bureaucracies, we conduct a large-scale investigation focused on the Chinese party-state. We theorize that, much like in democratic systems, bureaucratic selection is influenced by candidates' physical features because impressions of appearances provide intuitive (though not necessarily accurate) cues for forming beliefs about a person's innate characters, which are consequential for office-holding but inherently difficult to observe. However, we argue that important differences exist between selection through bureaucratic and democratic means: in elections, voters elect candidates to be their leaders, but bureaucratic decision-makers choose candidates who will *simultaneously* serve as a leader *and* a subordinate. An appointed official will be expected to not only be a competent leader for the lower-level agents but also a reliable and trustworthy subordinate to their superiors. The latter expectation can be particularly salient in centralized systems where concern about delegation and potential power abuses loom large. As a way to offset this potential threat, selectors are likely to favor candidates whose appearances convey an impression of a reliable and non-intimidating character.

To test this hypothesis, we collected over 21,000 official portraits and meeting and event photos for more than 4,000 party and government officials who held key leadership positions at the prefectural, provincial, and national level offices between 2000 and 2022. We manually rated a subset of these photos on four key traits—attractiveness, competence, trustworthiness, and aggressiveness— and used these human ratings to train a supervised machine-learning model capable of conducting automated appearance assessments at scale. Using machine-generated ratings, we examined how various facial traits correlate with two important career outcomes: (1) the highest political rank an official attained (promotion) and (2) whether an official was demoted or arrested

for criminal or disciplinary charges (purge), controlling for a host of other personal and career attributes.

Our empirical results show that the machine-generated ratings on the four perceptual traits are systematically associated with both positive and negative career outcomes of Chinese officials. Specifically, officials who are perceived to be more competent, trustworthy, and less aggressive are more likely to land on higher-level positions and face a lower likelihood of purge in their career. While the importance of competence is consistent with established findings about voter preferences in democratic elections (Castelli et al. 2009; Olivola and Todorov 2010a; Poutvaara, Jordahl, and Berggren 2009; Todorov et al. 2005), the strong effects of trustworthiness and non-aggressiveness on career success appear unique to the bureaucratic setting. We further find that appearing trustworthy and non-aggressive is particularly important for promotions to the more senior full-ministerial and national-level posts, compared to the relatively junior deputy-provincial ones, and the career penalty for appearing aggressive applies only to male candidates but not to female ones. These patterns are broadly consistent with the claim that, in a state of information asymmetry, facial traits are used by superiors in the Chinese bureaucracy to identify individuals with reliable characters and to filter out potential threats.

In addition to demonstrating that facial appearances matter for selection choices, we also assessed the substantive impact of facial traits relative to other factors. We estimated Random Forest models that included a host of officials' personal and professional attributes along with facial ratings to predict their career outcomes. We find that the predictive power of facial traits is at least comparable to, if not greater than, many conventional variables, such as an official's education, age, economic and fiscal performance, and political connections. Furthermore, we conducted a series of conjoint experiments with current Chinese government officials to gauge the relative magnitude of face-based preferences at the individual level. Our experiments presented subjects with pairs of hypothetical candidate profiles that include simulated facial photos, and asked subjects to select their preferred candidate for promotion. Consistent with the observational findings, government insiders exhibit strong preferences for candidate profiles with photos that are perceived to be more

competent, trustworthy, and less aggressive. The advantage conferred by a favorable appearance is comparable in magnitude to having a graduate degree or being a personal aide to a senior leader.

Our study engages with, and contributes to, several strands of literature. First and most directly, it speaks to a growing body of scholarship on the influence of facial impressions on social and political behavior. A wealth of social science scholarship has established that human faces are a rich source of stimuli for social cognition and trait inference (Dotsch et al. 2008; Freeman et al. 2009; Todorov 2017; Zebrowitz 1997). Individuals form rapid, sometimes unreflective, impressions based on facial appearances (Bar, Neta, and Linz 2006; Olivola and Todorov 2010a; Todorov, Pakrashi, and Oosterhof 2009), and these impressions can significantly influence real-world choices and actions (Benjamin and Shapiro 2009; Rule and Ambady 2008; Todorov et al. 2005). In politics, researchers have found that perceived attractiveness and competence strongly predict candidates' electoral performance across diverse settings. However, the role of facial appearances in non-electoral selection processes has received limited attention.[2] We contribute to this literature in two ways: First, we extend its scope to top-down selection in a civilian bureaucracy.[3] Our results show that while the preference for certain facial traits, such as competence, is consistent across institutions, others are unique to the mission and *modus operandi* of bureaucracies. This suggests that institutional and organizational contexts play an important role in shaping the characteristics of leaders who emerge. Second, we also make a methodological contribution by developing an automated approach to measure the average impressions of faces. Using a state-of-the-art computer vision algorithm, our method learns from human judgments and reproduces them with a decent level of accuracy. It can serve as a valuable tool for future facial analyses that involve a large number of subjects and numerous perceptual dimensions.

More broadly, our study also provides new evidence on the nature of the cognitive processes in organizational decision making. The human ability to read and interpret facial traits falls within

[2]The main exceptions are Mazur, Mazur, and Keating (1984) and Mueller and Mazur (1996), who show that West Point cadets with a more socially dominant look achieved higher military ranks both at graduation and later in their careers.

[3]Top-down selection is likely to matter even in electoral settings, as candidates often have to be winnowed by party leaders and elites before facing voters (Gallagher and Marsh 1988). The preferences and considerations that shape intra-party selection may be similar to those observed in a bureaucratic context.

what Polanyi (1967) calls the tacit domain of knowledge—skills and expertise that are learned through direct experience and practice but are difficult to express or explain verbally. The modern dual process model in social psychology similarly proposes that the brain operates through two systems: System I is implicit, fast, unconscious, and intuition-based, whereas System II is explicit, slower, conscious, and logical (Epstein 1994; Khaneman 2003; Stanovich and West 2000). While it is now well established that the tacit, System I-style processing constitutes a significant part of human cognition, existing social science theories about institutions have still maintained a largely rationalist orientation, emphasizing actors' strategic calculations based on System II processes.[4] By showing that perceptions and impressions influence high-stake personnel decisions in a political organization known for its rigorous and meticulous cadre selection procedures, we provide systematic evidence that intuitive judgments may underlie many seemingly rational organizational decisions.

Finally, we contribute to the literature on political selection in contemporary China. Existing scholarship on this topic typically follows one of two paradigms (Manion 2023): the performance paradigm, which emphasizes the importance of formal institutions and quantitative metrics (Li and Zhou 2005; Liu 2023; Xu 2011; Yao and Zhang 2015), and the patronage paradigm, which underscores the role of personal connections with higher-level decision-makers (Keller 2016; Nathan 1973; Shih, Adolph, and Liu 2012). We move beyond these two paradigms by highlighting a different set of factors rooted in perceptions of candidates' appearances. Recent empirical studies have provided suggestive evidence that physical appearance matters in China's selection process, but they have tended to focus either on the difference in facial features between elites and non-elites (Wang, Li, and Praino 2024; Wong and Zeng 2017), or on physical attractiveness as the sole predictor for political mobility (Ling, Luo, and She 2019). We build on this emerging body of research and extend it by adopting a multidimensional approach to measuring facial traits and examining the

[4]Early organizational theorists, such as Barnard (1936) and Simon (1947), did distinguish between "logical" and "non-logical" processes in executive decision making. According to Barnard (1936), for example, the logical process refers to conscious reasoning that can be expressed in words or symbols, whereas the non-logical process refers to the rapid, spontaneous judgment generated from intuitive responses rather than analytical reflection.

influence of appearance on promotion across multiple administrative levels.[5] Our findings suggest that the perceptions of appearance can be as consequential as performance metrics or patronage ties in shaping officials' careers. However, rather than rewarding attractiveness per se, the system appears to place greater weight on traits such as trustworthiness and non-aggressiveness. These findings point to alternative way of conceptualizing selection in the Chinese bureaucracy: Instead of viewing it solely as a meritocratic "tournament" where candidates compete on objective qualifications (Li and Zhou 2005; Xu 2011), the process may sometimes resemble a "beauty contest," in which perceptions of appearance also play a significant role in determining success.

## The Perceptual Dimension of Bureaucratic Selection

The staffing and appointment practices of government bureaucracy have long been the subject of interest for a vast, multi-generational body of scholarship. According to the classical model proposed by Weber (1921 [1978]), bureaucracy is a rational and efficient form of organization that embodies the legal-rational authority characteristics of modernity. Recruitment and selection processes within bureaucracies are supposed to be objective, rule-based, and meritocratic, prioritizing impersonal criteria such as seniority, performance, and technical qualifications over personal preferences and biases (Merton 1940; Rauch and Evans 2000).

While the Weberian model is highly influential as an ideal type, later research has challenged many of its core assumptions about bureaucratic organizations. In particular, critics have argued that real-world bureaucratic organizations rarely uphold the rational and meritocratic standards in candidate selection as strictly as Weber has postulated (Downs 1967; Rudolph and Rudolph 1979). One reason is that bureaucratic performance is inherently difficult to measure: Government

[5]For instance, Ling, Luo, and She (2019) find that physical attractiveness is positively associated with local leaders' promotions, but they do not account for other perceptual traits that may be correlated with attractiveness. Wang, Li, and Praino (2024) show that ordinary citizens can identify political leaders with above-chance accuracy and that local political leaders exhibit distinct facial characteristics compared to non-political leaders. However, they do not examine how appearance-related traits affect career mobility *within* the government. Finally, Wong and Zeng (2017) find that perceived facial competence predicts the election of rural deputies to local People's Congresses but not the promotions of government officials.

work is usually collective in nature and considerable task heterogeneity exists across agencies (Thiel and Leeuw 2002; Wilson 1989). When individual performance cannot be precisely gauged and compared, the personal discretion of senior decision-makers plays a significant role in the evaluation of agents. Yet, given that the superiors often lack sufficient time, energy, or attention for a full appraisal of every subordinate (Simon 1947), crucial to an official's career advancement is not only to achieve good performance but also to become "visible" to those who can influence their appointment within the organization (Moore and Trout 1978). This means both competing for the attention of higher-level decision-makers and leaving a favorable impression upon them within a limited period of interaction (Janvry et al. 2023). In line with this visibility-centered perspective, studies have shown that social connections and shared group identities with superiors can facilitate promotion by raising candidates' profile salience and front-loading favorable impressions with decision-makers (Grindle 2012; Rudolph and Rudolph 1979).

Among the various factors that can influence a candidate's visibility, facial appearance may be particularly important.[6] Faces are among the richest and most powerful tools for humans' social communication and cognition (Hassin and Trope 2000; Jack and Schyns 2015; Todorov 2017; Zebrowitz 1997). They provide a wealth of stimuli for observers to infer traits such as physical attractiveness (Rhodes 2006), personality traits (Todorov et al. 2015; Rule and Ambady 2011), and skills and competence (Eisenbruch et al. 2024; Todorov et al. 2005). Facial stimuli are processed through a distributed neural system that involves multiple regions located primarily in the right hemisphere of the brain (Haxby, Hoffman, and Gobbini 2000).[7] Responses to facial stimuli are spontaneous, effortless, and unreflective, often occurring within fractions of a second (Olivola

[6]The idea that someone's innate traits can be inferred from their faces has a long history in many civilizations. In the 18th and 19th century, the pseudoscience of physiognomy enjoyed tremendous popularity in European intellectual circles. Figures like Johann Caspar Lavater, Francis Galton, and Cesare Lombroso published numerous treatises—now largely discredited—linking facial appearances to character traits and even criminal predispositions. For reviews of the history of physiognomy, see Brandt (1980) and Todorov (2017)

[7]Research in neuropsychology has identified two primary pathways for face processing: The ventral pathway, which predominantly processes invariant structural and surface properties of faces, includes the occipital face area (OFA) and the fusiform face area (FFA). The dorsal pathway, which is more responsive to dynamic facial elements such as expressions, gaze, and mouth movements, consists of face-selective areas in the posterior superior temporal sulcus (pSTS), the anterior superior temporal sulcus (aSTS), and the inferior frontal gyrus (IFG). For a review of relevant work, see Duchaine and Yovel (2015).

and Todorov 2010a; Willis and Todorov 2006). There is also some evidence that people across different cultures and nationalities share considerable agreement over the impressions they make on the same face (Rule et al. 2010; Zebrowitz, Montepare, and Lee 1993).

Existing studies suggest that the impressions that individuals form about a face can be clustered into three broad dimensions: (1) perceived capability (e.g., competence, dominance), (2) perceived social warmth (e.g., trustworthiness, agreeableness), and (3) youthful-attractiveness (Fiske, Cuddy, and Glick 2007; Oosterhof and Todorov 2008; Sutherland et al. 2013; Sutherland et al. 2015). In electoral contexts, research has found that voters systematically prefer candidates whose faces look competent and attractive, but not necessarily those who score highly on the warmth dimension (Joo, Steen, and Zhu 2015; Todorov et al. 2005). In the context of a bureaucracy, however, we may expect the preferences for facial traits in bureaucratic selection to differ due to the distinct goals and institutions governing the selection decisions.

Specifically, we argue that a key feature that sets bureaucratic selection apart from elections is that the goal of bureaucratic selection is to select *both* a leader and a follower. Unlike elected politicians, who in theory serve no "bosses" other than their constituency as a whole, bureaucratic officials must fulfill dual roles: they need to be a competent leader for their subordinates and a reliable and loyal subordinate to their superiors. This duality has important implications for the types of appearances favored in bureaucracy. When selecting leaders, decision-makers often prioritize traits associated with competence, self-confidence, and social dominance (Judge et al. 2002; Lord, Vader, and Alliger 1986), as these qualities signal an individual's ability to give directions in times of uncertainty and enforce compliance in challenging multi-person collective actions (Laustsen and Petersen 2015). However, traits indicative of power and dominance may be less desirable in selecting subordinates, whose prototypical qualities include agreeableness, loyalty, and reliability (Sy 2010). In hierarchical organizations, where the problem of delegation is pervasive and interpersonal trust is critical, superiors may especially value those candidates who can dutifully carry out orders given to them (Moe 2012). As one moves up on the hierarchy of power, the promoted will be entrusted with a level of power that could be used to harm those who appointed them. It

is thus unsurprising that loyalty, reliability, and non-aggressiveness become highly sought-after qualities at the top of political and administrative hierarchies, sometimes even at the expense of competence (Egorov and Sonin 2011; Wagner 2011). Since these qualities cannot be observed directly, decision-makers often have to rely, consciously or unconsciously, on visual heuristics, leading them to favor candidates whose facial features project a trustworthy and non-intimidating image.

Taken together, the preceding discussion suggests while there are important reasons to believe that visual cues such as facial appearances play a significant role in the ostensibly rational process of bureaucratic selection, the unique concerns and priorities faced by bureaucratic decision-makers may lead them to prefer a somewhat different set of appearance traits than those valued by voters in elections. Specifically, we hypothesize that *preference will be given not only to an image of competence, which is typically associated with effective leadership, but also to trustworthiness and non-aggressiveness, which convey the impression of a loyal and reliable subordinate*.

## Appearance-based Cues and Political Selection in China

The question of what shapes the selection of party and government officials within the Chinese party state is one of the most central issues in contemporary Chinese politics research. As summarized by Manion (2023), most of the existing studies follow one of two dominant paradigms. The performance paradigm views selection as a highly institutionalized process guided by meritocratic criteria. Influential studies have shown that factors such as economic and fiscal revenue growth are positively correlated with promotion prospects of local officials (Li and Zhou 2005; Yao and Zhang 2015). By contrast, the patronage paradigm challenges the binding nature of formal rules, emphasizing instead the role of private political and financial interests in shaping promotion decisions. Researchers following this paradigm argue that informal patron-client relations with higher-level leaders, or factional ties, are the primary determinants of an official's political fortune within the system (Huang 2000; Keller 2016; Nathan 1973; Shih, Adolph, and Liu 2012). More recently,

some studies have sought to reconcile these two paradigms by exploring how public and private considerations may coexist. For instance, Landry, Lü, and Duan (2018) propose that performance- and patronage-based factors may operate at different levels of government, with lower-level promotions favoring performance and higher-level promotions prioritizing patronage. Jia, Kudamatsu, and Seim (2015), moreover, suggest that the two may be complementary, as the reward for performance is often greater for connected officials than unconnected ones.

While these two paradigms offer seemingly contradictory explanations for the selection practice in the Chinese bureaucracy, we argue that they nonetheless share a crucial implicit assumption: that is, selection decisions are based on a rational and informed evaluation of candidates' "real" strengths and weaknesses—whether they are professional achievements or factional affiliations. This assumption is rooted in both the long-standing perception of communist parties as a highly capable Leninist organizational machinery (Selznick 1952) and the more recently developed myth about the omniscient record-keeping power of the Organization Department (McGregor 2010). Viewed in this light, the Chinese party state would be the least likely case where appearance-based selection practice would matter.

However, a closer examination of selection practices within the Chinese system reveals a more complex reality. In most cases, the selection process operates neither like a precise, perfectly synchronized machine nor like a highly strategic factional game played by a few. Instead, it more closely resembles an athletic draft, where the final choice is shaped not only by a candidate's objective qualifications but also by subjective impressions from a diverse audience of stakeholders.[8] As illustrated in Figure 1, a typical promotion process involves multiple stages and incorporates input from numerous individuals. At the nomination stage, for example, a key step is democratic recommendation (民主推荐), in which the higher-level party committees invite suggestions and feedback on potential candidates from leading members of lower-level party, government, judiciary, and mass organizations. The scope of consultation is extensive: For a prefecture-level appointment, it sometimes solicits written input from more than 200 individuals and conducts interviews

[8]For a related discussion on the performative aspect of public-facing governance, see Ding (2022)

with over 100 (Zeng 2015). Once the nominations are made, the OD dispatches an assessment team to inspect and vet each candidate's career and personal background. According to the official guidelines, candidates are to be evaluated by five general criteria: morality 德, competence 能, diligence 勤, achievements 绩, and probity 廉. These qualities, however, are inherently difficult to observe and quantify, as they pertain to a person's innate character and integrity. Consequently, evaluations in these domains have to rely not only on objective professional records but also on subjective appraisals from face-to-face interviews (and, in some cases, polling) with the candidate's supervisors, co-workers, and subordinates. After the assessment is completed, the OD team produces a detailed report, which is reviewed and deliberated upon by the higher-level party committee. The final decision of promotion is again a collective one, requiring a two-thirds quorum of the standing committee and a majority vote. Once a decision is reached, it is publicly announced for a period of 7 to 15 days, during which concerned individuals may raise objections against the appointment, and additional legislative approval is needed if appointments are for state positions.

As this lengthy process reveals, earning a promotion within the Chinese bureaucracy requires candidates to not only please their immediate superiors, but also be able to impress a broad group of evaluators and potential veto players. Since not all evaluators have the opportunity or energy to develop long and close relationships with candidates, a significant part of their judgment is likely based on limited interactions or indirect sources. This creates room for salient features, such as facial appearance, to influence promotion preferences. If a candidate's appearance convincingly conveys certain desirable qualities, they may have a better chance of making a favorable impression on decision-makers within a short period of time. Appearances not only directly influence perceptions of an official's personal qualities, but can also indirectly influence their career by affecting the evaluations of work performance or the configurations of social networks. For example, while successful policy projects are often the result of collective efforts, officials with a more competent appearance may receive disproportionate credit for the success (Fan et al. 2018). Additionally, those who appear more trustworthy and non-threatening may be more likely to be recruited into the inner circles of senior leaders, designated as close aides, or even groomed as successors.

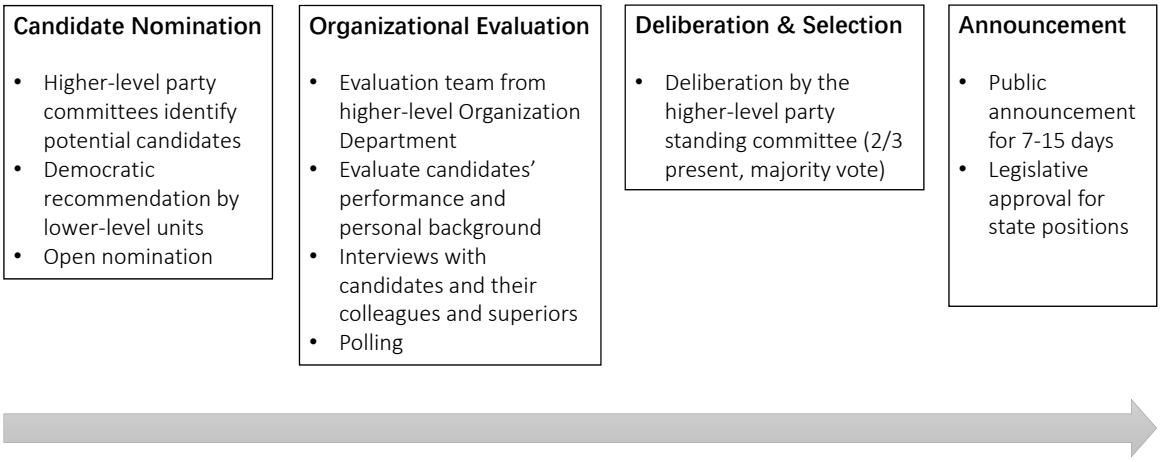
While direct, systematic evidence remains limited, several anecdotes and recent studies suggest that (1) decision-makers take into account candidates' personalities when making promotion decisions, and (2) facial appearance is viewed as a relevant factor for career success by government insiders. In terms of personality preferences, there is evidence that warmth-related traits are favored in the selection process. For example, Chen Yun, one of the top leaders in the 1980s known for his expertise in personnel affairs, once wrote that cadres selected for promotion must not only be competent and morally upright, but also "neither prickly nor hard to get along with" (Chen 1995, 293–294). A recent study using the OD's character assessments found that provincial leaders who are perceived as having a more collegial leadership style and humble personality are less likely to face corruption investigations (Jiang and Luo 2021). Moreover, indirect evidence suggests that facial appearance serves as an implicit yet significant criterion of assessment used by government insiders. A nationally representative survey on county-level civil servants conducted in 2005 found that approximately 28.3% of respondents believed "face reading" could predict a person's future success—a proportion higher than observed among the general public (26.7%), even though civil servants are generally much better educated and possess more scientific knowledge than the average citizen.

# Data and Methods

## Collecting and Pre-processing Officials' Facial Images

To systematically evaluate the effect of facial appearance on bureaucratic career mobility in China, we collected and analyzed facial images from a large sample of mid- and senior-level party-state officials. Our dataset includes all individuals who served at least one of the following key prefecture, provincial, or national positions between 2000 and 2022: (1) city mayors and party secretaries, (2) provincial standing committee members, (3) provincial governors, party secretaries, and ministers, and (4) Politburo members, as well as chairmen and vice-chairmen of national assemblies (the National People's Congress and the Chinese People's Political Consultative

Figure 1: Formal Selection Procedures in the Chinese Bureaucracy




***Note:*** This figure provides a stylized illustration of the typical promotion process within the Chinese bureaucracy. Information is based on the CCP's *Regulations on the Selection and Appointment of Party and Government Leading Cadres* (2014 edition).

Conference).[9] The full sample consists of 5,060 unique individuals. For each official, we gathered both their official frontal headshots and photos taken during political meetings or other public events. We selected only photos in which the subject faces forward with a neutral expression.[10] We included meeting photos as a way to address the issue of reverse causality: higher-ranking individuals' official headshots might exhibit more desirable traits because they are more likely to be extensively edited compared to those of lower-ranking officials. In contrast, meeting and event photos usually involve less staging and editing, making them more likely to capture the true appearance of officials. The final database includes 7,484 frontal headshots and 13,783 meeting and event photos. Each photo was cropped to include only the facial region above the neck, and the original backgrounds were replaced with a white backdrop. All photos were standardized to a resolution of 413 x 579 pixels.

[9]These positions constitute the vast majority of the senior leadership posts in the Chinese system. The only significant group of officials not currently included in our sample is vice-provincial governors and vice-ministers in central ministries. We omitted them because of limited availability of detailed appointments and personal information for this group.

[10]Photos were collected from a variety of online sources, including Wikipedia, Baidu Encyclopedia, and news platforms that regularly cover the activities of central and local officials.

## Predicting Facial Traits with Deep Learning

A central methodological challenge to our empirical analysis is obtaining subjective assessments of facial appearances for a large number of officials. Currently, the most common approach is to recruit human respondents to rate photos of all political figures being studied and generate aggregate ratings based on the responses. This method encounters two problems when being applied to the Chinese context. The first issue is familiarity: Higher-ranking officials are typically more publicly recognizable than lower-level figures, and this differential publicity directly confounds the relationship between appearances and political ranks. Raters may associate certain leadership-related traits with a higher-ranking figure because they have seen this person in a leadership position. The second issue is fatigue: Given the sheer size of our sample and the multiple dimensions that we are interested in, we would either need to recruit a very large number of respondents or impose an excessive workload to a team of a feasible size. In the latter case, mental and vision exhaustion from long and repetitive surveys could significantly compromise response quality, especially for a perceptual task like ours (Jeong et al. 2023; Hirao et al. 2021).

To address these challenges, we adopted an alternative, deep-learning method to perform the measurement task at scale. Specifically, we use Convolutional Neural Networks (CNNs) to "learn" from a relatively small set of human ratings on officials' facial appearances and then apply the same standard to rate the remaining photos in the dataset.[11] Originally conceptualized by LeCun et al. (1989), CNNs are known for their efficient learning capabilities and have demonstrated exceptional performance in computer vision tasks. The model we use in this study is ResNeXt-50, an advanced algorithm designed for image recognition and feature extraction.[12] ResNeXt-50 builds on the traditional CNN architecture by incorporating methods such as residual connections and grouped convolutions, which enable the training of very deep networks and efficient parallel

[11] A key assumption underlying this approach is that there is a reasonable degree of consensus in how people perceive and evaluate facial features. This assumption is supported by studies showing considerable agreement in facial perceptions across different cultures, race, and nationalities (Albright et al. 1997; Lawson et al. 2010; Rule et al. 2010).

[12] We also tested other commonly used algorithms, including AlexNet, DenseNet, and ResNet. ResNeXt-50 was selected for its high performance, lightweight architecture, and fast convergence. For detailed comparisons, see Table B.3 in the Online Appendix.

processing of granular details on the images (Lin, Liang, and Jin 2019). Below, we outline the workflow of our methodology, with a schematic overview provided in Figure B.1 of the Online Appendix.

**Human Rating for Facial Traits** We began by generating a training dataset of human-rated photos. We recruited 199 raters to evaluate 2,500 photos randomly selected frontal headshots of city- and provincial-level officials.[13] The raters were hired online from a company that specializes in data annotation services. The modal rater is female (54.8%), between the age of 30 and 40 (55%), and has an education level of high school or lower (49.2%).[14] Each rater was given 100 randomly selected photos from the training set, along with 25 fixed photos that all raters were required to rate. Raters were instructed to evaluate each photo on four facial traits—attractiveness, trustworthiness, intelligence, and aggressiveness—using a scale from 1 to 5. We focus on these four traits because they are both important considerations in leadership selection (Vugt and Grabo 2015) and fundamental dimensions in social cognition and interpersonal evaluation (Fiske, Cuddy, and Glick 2007; Oosterhof and Todorov 2008; Sutherland et al. 2013). Trustworthiness and (the inverse of) aggressiveness reflect evaluations on the social warmth dimension, which relates to perceived intent (good or ill); competence captures an evaluation of one's ability (high or low) (Fiske, Cuddy, and Glick 2007); and attractiveness is linked to perceptions of youthfulness and mate preferences (Sutherland et al. 2013).[15] After excluding photos rated by fewer than three raters, our training set has a total of 2,403 photos. Inter-rater reliability is reasonably high across

[13]The training dataset focuses only on the official headshots of city- and provincial-level figures for two reasons. First, profile pictures are more uniform in style and quality, making them more suitable for consistent labeling. This uniformity ensures that the ratings are not influenced by variations in image quality or style, such as tilted faces or differing lighting conditions, which are more common in meeting photos. Second, lower-level officials are less likely to be recognized by raters, thereby minimizing the familiarity bias. We empirically validated this assumption in a supplementary recognition survey in Appendix B.2. The survey results suggest that the correct identification rates for officials at city (1.4%), deputy-provincial (0.7%), and provincial level (1.5%) are generally extremely low and statistically indistinguishable from one another. The rates are somewhat higher for national leaders (2.4% for non-Politburo Standing Committee members and 3.6% for PSC members).

[14]For additional details about the raters, see Appendix B.1

[15]A somewhat different classification of perceptual dimensions has been proposed by Oosterhof and Todorov (2008). Using a principal component analysis of trait judgments on emotionally neutral faces, they identify two orthogonal dimensions—valence and dominance. In our study, trustworthiness, and to some extent attractiveness, map to the valence dimension, while aggressiveness corresponds to the dominance dimension.

all four traits, suggesting that there is considerable agreement among raters in their assessments of faces (Table B.1). This manual labeling process provides the ground truth data for training the deep learning model.

**Model Training** We fed pixel-level information of manually labeled photos into the pre-trained ResNeXt-50 model after some modest preprocessing.[16] The ResNeXt-50 model consists of multiple layers designed to progressively refine and abstract the input data. The initial layers extracted basic image features—such as edges, textures, and shapes—through convolution and pooling operations. These features were encoded in the feature maps, which are two-dimensional arrays that capture the spatial and semantic information of an image at various levels of abstraction.[17] The model then used bottleneck layers to refine these extracted features. After the convolutional and bottleneck layers, the feature maps were processed by an average pooling layer, which reduced the spatial dimensions of the feature maps and produced a compact representation of the facial features across the entire image. Finally, the output layer used a linear function to transform this compact representation into a continuous score for each facial trait, ranging from 1 to 5.

**Diagnostics and Validation** We evaluated the performance of the trained model in several ways. First, we assessed its predictive accuracy using standard metrics, including the Root Mean Squared Error (RMSE). The Root Mean Squared Error (RMSE) ranged from 0.45 to 0.69 on a 1–5 scale, with slightly higher errors for attractiveness compared to the other traits.[18] Second, we conducted a 10-fold cross-validation on the human-labeled image set. This exercise involves partitioning the training data into ten equal folds and iteratively using a model trained on nine folds to generate

[16] We applied two data pre-processing functions, `ToTensor`, which scales the pixel values from the $[0, 255]$ range to the $[0, 1]$ range, and `Normalize`, which standardizes the pixel value distribution to zero mean and unit variance. We did not use data augmentation techniques, such as rotation or flipping, for two reasons. First, our data are standardized headshots that follow strict conventions, including frontal facial orientation and fixed backgrounds. Augmentations like rotation or flipping could distort the semantic integrity of the images, leading the model to learn irrelevant features that compromise prediction accuracy. Second, the relatively small sample size increases the model's susceptibility to overfitting, and augmentations deviating from the original data distribution may exacerbate this issue.

[17] Spatial information refers to the location and arrangement of features within the image, and semantic information is about the meaning or interpretation of detected features.

[18] See Table B.3 in the Online Appendix for additional diagnostic metrics, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE).

out-of-sample predictions for the remaining one.[19] Figure 2 plots the out-of-sample predictions against the actual human ratings. We see positive and statistically significant correlations across all four traits ($p < 0.001$ for all). As an additional check, we recruited a new group of raters to rate another sample of photos ($n = 160$) and compared their ratings with the model's original predictions. Again, we find strong correlations between human and machine ratings (Pearson's $r > 0.39$ for all traits, see Appendix B.4 for details).[20]

As a visual validation of the output, Figure 3 presents synthetic images generated by morphing actual photos within various rating percentiles. For each trait dimension (row), the images on the far left and right represent the averages of the 15 highest-rated and the 15 lowest-rated photos for that trait, respectively. The images in the middle are averages from 15 photos randomly drawn from the percentile interval indicated on the top (60th to 80th, 40th to 60th, and 20th to 40th). We can see that subtle yet notable differences exist in appearances across the rating spectrum for both female and male officials.[21] Faces rated as more attractive tend to appear younger, more symmetrical, and have a lower width-to-height ratio. Faces associated with higher competence ratings have longer noses and more pronounced chins, while rounder and softer facial features are typical of faces rated high on trustworthiness. Aggressive-looking faces often exhibit angular edges, narrower eyes, and downward-turned mouth corners.[22]

## Empirical Specification

Our main estimation framework uses a fixed-effects model with the following specification:

[19] This procedure ensures that every image's predicted score was generated by a model that had not been exposed to it during training, thus providing an unbiased, out-of-sample estimate of model performance.

[20] Since human raters have a well-documented tendency to give scores in the middle (Alfertshofer et al. 2024), we used a stratified random sampling strategy that slightly over-sampled the top and bottom distribution of the photos to ensure that sufficient variation exists in human ratings.

[21] The visual differences between faces at the opposing ends of each trait spectrum are generally smaller in synthetic photos than real photos, because averaging tends to eliminate idiosyncratic features of individual images, resulting in a convergence in appearances. For a comparison of actual photos with extreme ratings, see Figure B.9 in the Online Appendix.

[22] In Figure B.8 in the Online Appendix, we report bivariate correlations between the machine-based ratings and a host of physical features of faces (e.g., face length, face width, eye size, etc).

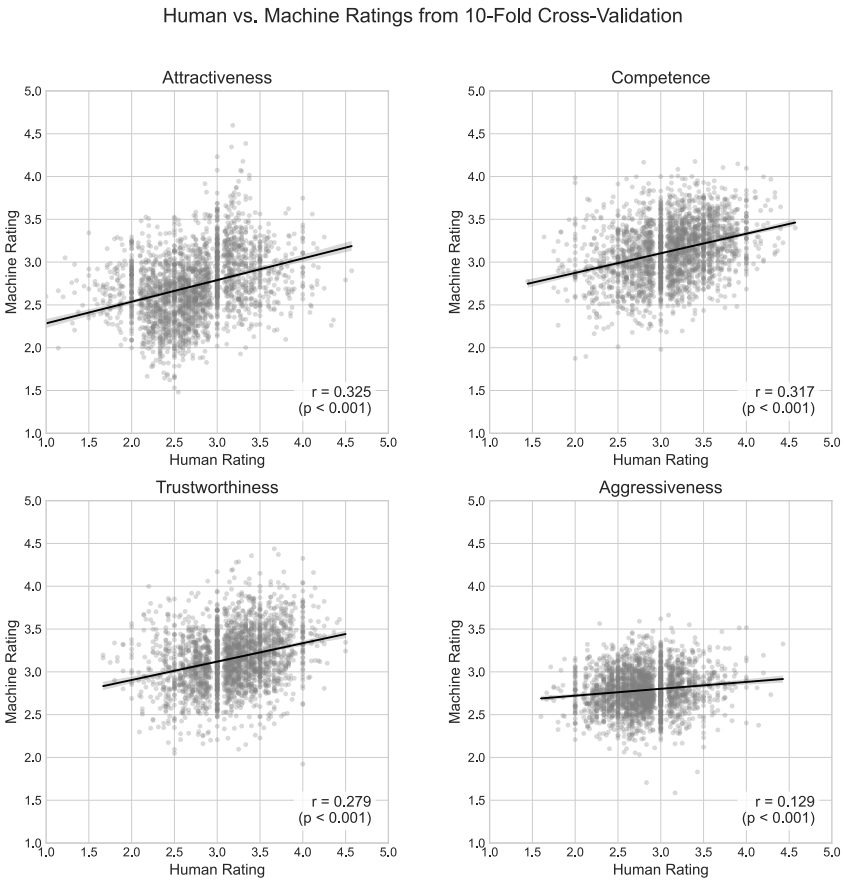
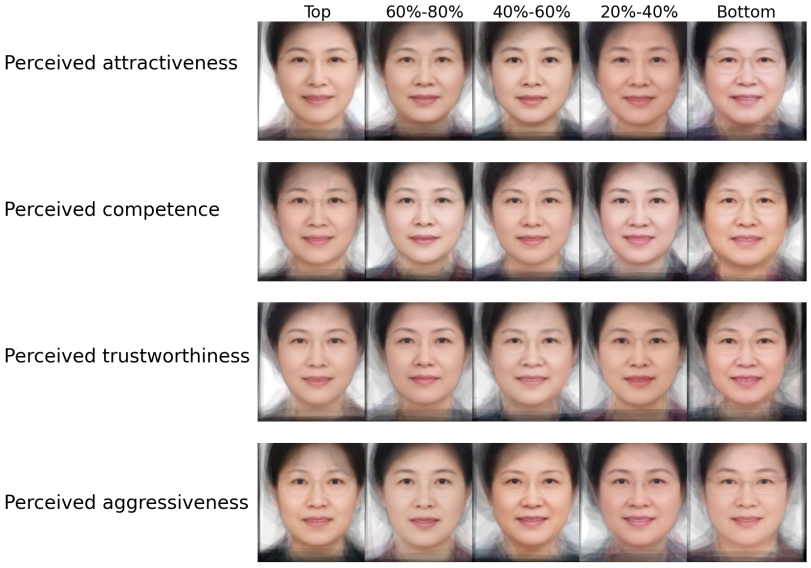


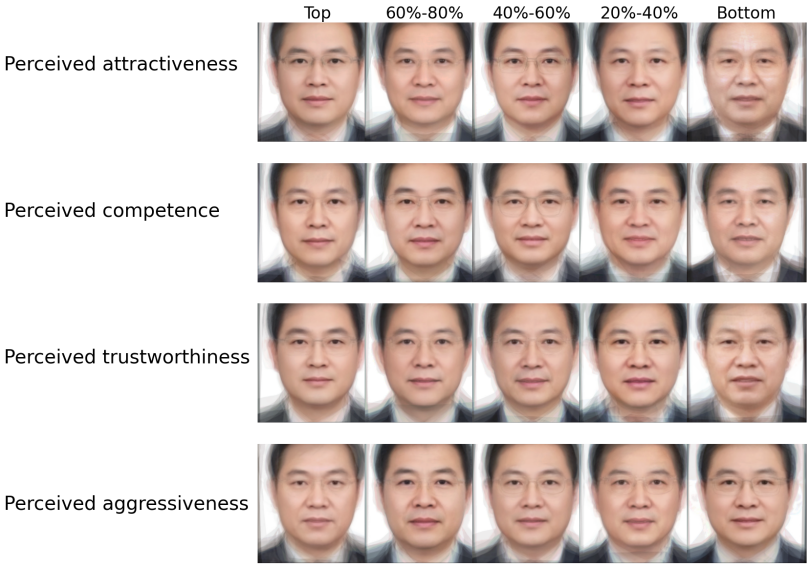

Figure 2: Human vs. Machine Ratings from 10-Fold Cross-Validation

*Notes*: The figure displays scatter plots of out-of-sample machine predictions against human ratings for all images in the training dataset, based on a 10-fold cross-validation procedure. Each point represents a single image. The solid black line is the linear best-fit line. The text box in each subplot reports the Pearson correlation coefficient (r) and its p-value.

Figure 3: Average Faces by Facial Rating and Gender




Female Officials




Male Officials

***Note:*** This figure displays synthetic faces generated by morphing actual photos of officials within varying rating percentiles. The images on the far left and right are created by averaging the 15 highest-rated and the 15 lowest-rated photos for each trait, respectively. The images in between are averages based on 15 photos randomly drawn from the percentile intervals indicated at the top.

$$Y_i = \delta^k \text{Facial trait rating}_i^k + \mathbf{X}_i\boldsymbol{\beta} + \eta_i + \gamma_i + \epsilon_i, \quad (1)$$

where $i$ indexes an individual official. We use two main dependent variables: *Maximum rank*, which measures the highest rank an official attained by 2022, and *Purge*, a binary indicator that takes the value 1 if an official was ever demoted, arrested, or subjected to party disciplinary sanctions. The main independent variables, *Facial trait rating*$^k$, denote the machine-generated facial ratings for dimension $k$. $\mathbf{X}$ is a vector of covariates for an official's basic demographic characteristics, such as gender, ethnicity, and educational level. We also include in the baseline model fixed effects for an official's prefecture of birth ($\eta_i$) and year of birth ($\gamma_i$), as officials from different regions may exhibit distinct facial features and have different career potential due to variations in regional political importance. Likewise, officials born into different cohorts may adopt different styles and face disparate prospects for advancement at any given year. These fixed effects help account for hometown- and generation-specific factors that may confound the effect of appearance on career outcomes.

# Results

## Baseline Results

The baseline results are presented in Table 1. Models 1 to 5 report the estimated associations between facial trait ratings and the highest political rank attained. In Models 1 through 4, we examine each of the four facial rating variables individually. All four facial variables are significantly correlated with officials' political rank, suggesting that systematic differences in facial appearances do exist between officials at different ranks. In Model 5, we include all four traits in a single regression to account for potential correlations between facial ratings across different dimensions. Consistent with the robust findings from the electoral context (Castelli et al. 2009; Olivola and Todorov 2010a; Poutvaara, Jordahl, and Berggren 2009; Todorov et al. 2005), perceived com-

petence shows a positive and significant association with higher political ranks in bureaucratic selection as well. The coefficient estimate suggests that a one standard deviation increase in an official's perceived competence is associated with a 0.062-unit (or 7.4% of a standard deviation) increase in the expected maximum rank.

In addition to competence, the two warmth-related traits—perceived trustworthiness and (non-)aggressiveness—are also significantly associated with officials' career success. A one standard deviation increase in trustworthiness corresponds to a 0.051-unit increase in expected rank (6.1% of a standard deviation), while a one standard deviation increase in aggressiveness decreases the expected rank by 0.038 units (4.6% of a standard deviation). The significant effect of trustworthiness contrasts with its largely null impact in electoral settings (e.g., Berggren, Jordahl, and Poutvaara 2010; Mattes et al. 2010; Rosenberg et al. 1986; Rule and Ambady 2008; Todorov et al. 2005), but aligns with our theoretical expectations: trustworthiness and a non-threatening personality are particularly valued in bureaucratic systems, and decision-makers often rely on facial cues to identify candidates who possess these qualities.

The effect of perceived attractiveness, meanwhile, shows a different pattern. When included individually, attractiveness has a positive and statistically significant association with official rank. However, the coefficient becomes non-significant when other facial traits are included in the regression. This diverges from findings in electoral contexts, where attractiveness has a positive and independently strong influence on vote share (e.g., Banducci et al. 2008; Berggren, Jordahl, and Poutvaara 2010; Budesheim and DePaola 1994; Sigelman, Sigelman, and Fowler 1987). The lack of a direct effect for attractiveness may be explained by the fact that bureaucratic selection typically involves far fewer public-facing activities than elections. In some cases, an excessively attractive physical appearance might even be a liability, as it could suggest widespread popular support outside the bureaucracy or provoke jealousy among decision-makers (who are usually older and less physically attractive). Instead, bureaucratic systems may value attractiveness only insofar as it signals other desirable traits, such as competence or trustworthiness, and may prefer a "quieter" form of beauty—good-looking, but not overwhelming or imposing.

Models 6 through 10 report the results for negative career outcomes. Following a similar order, we first estimate each facial trait variable individually and then include them all in a single regression. When analyzed individually, perceived attractiveness, competence, and trustworthiness are each negatively and significantly associated with the likelihood of an official being purged. The estimate for perceived aggressiveness is positive but non-significant. The signs of the coefficients are the exact opposite of those observed for promotions. When all facial traits are included, the coefficient for perceived attractiveness once again cease to be significant, but those for perceived competence and trustworthiness remain strong.

Table 1: Baseline Results: Promotion and Purge

| | DV: Promotion | | | | | DV: Purge | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Perceived attractiveness | 0.262** | | | | 0.057 | -0.073** | | | | -0.026 |
| | (0.056) | | | | (0.060) | (0.024) | | | | (0.028) |
| Perceived competence | | 0.426** | | | 0.284** | | -0.098** | | | -0.065** |
| | | (0.064) | | | (0.069) | | (0.021) | | | (0.025) |
| Perceived trustworthiness | | | 0.411** | | 0.262** | | | -0.096** | | -0.059* |
| | | | (0.066) | | (0.072) | | | (0.026) | | (0.028) |
| Perceived aggressiveness | | | | -0.296** | -0.241** | | | | 0.011 | -0.002 |
| | | | | (0.068) | (0.067) | | | | (0.026) | (0.027) |
| Female | 0.218** | 0.233** | 0.227** | 0.220** | 0.217** | -0.034* | -0.039* | -0.037* | -0.038* | -0.037* |
| | (0.051) | (0.051) | (0.051) | (0.052) | (0.051) | (0.016) | (0.016) | (0.016) | (0.016) | (0.015) |
| Han ethnicity | -0.056 | -0.050 | -0.055 | -0.056 | -0.052 | -0.006 | -0.008 | -0.006 | -0.006 | -0.007 |
| | (0.057) | (0.055) | (0.055) | (0.055) | (0.055) | (0.020) | (0.021) | (0.021) | (0.021) | (0.021) |
| College degree | 0.490* | 0.490** | 0.524** | 0.519** | 0.519** | 0.130** | 0.129** | 0.121** | 0.126** | 0.126** |
| | (0.193) | (0.189) | (0.190) | (0.189) | (0.188) | (0.028) | (0.029) | (0.027) | (0.028) | (0.028) |
| Graduate degree | 0.641** | 0.640** | 0.679** | 0.675** | 0.669** | 0.147** | 0.146** | 0.137** | 0.141** | 0.143** |
| | (0.200) | (0.196) | (0.198) | (0.196) | (0.195) | (0.027) | (0.028) | (0.027) | (0.028) | (0.028) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.25 | 0.26 | 0.26 | 0.25 | 0.26 | 0.12 | 0.13 | 0.13 | 0.12 | 0.13 |
| Observations | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 |

***Note***: This table presents the estimated effects of facial trait ratings on two sets of career outcomes: promotion and purge. The dependent variable for Models 1 to 5 is the maximum political rank an official attained in his/her career as of 2022, and the dependent variable for Models 6 through 10 is a binary indicator for whether an official was ever demoted or arrested due to criminal charges or violations of party disciplines by 2022. FE = fixed effects.
Standard errors clustered at the birth city level are reported in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

## Results by Promotion Steps

The baseline analysis uses a pooled sample of officials from four different ranks. However, since senior and junior positions often entail different levels of power and types of responsibilities, the facial traits deemed desirable for candidates may differ across ranks. To explore this potential heterogeneity, we constructed three subsamples, each containing all officials who served at a single rank (i.e., prefecture, deputy-provincial, and full-provincial). For each subsample, we created a binary indicator for whether an official ever advanced to the next higher rank and used it as the dependent variable to replicate the baseline analysis.[23]

As shown in Table 2, the relative influence of different facial traits shifts notably across promotion steps: At the lowest step (from prefecture to deputy provincial), perceived competence has the largest impact on promotion among all facial traits: a one-standard-deviation increase in perceived competence raises the probability of promotion for prefecture-level officials by 1.8 percentage points. This emphasis on competence at this level likely reflects the need to select leaders who are capable of managing complex, day-to-day governing tasks. The finding is consistent with Landry, Lü, and Duan (2018), who show that substantive performance plays a more consequential role in promotions at lower levels. Moving to the next higher step (deputy provincial to full provincial), the effect of competence remains comparable (2.4 percentage points for a one-standard-deviation increase), but perceived trustworthiness and non-aggressiveness grow substantially more important. A one-standard-deviation increase in trustworthiness raises promotion probability by 3.7 percentage points, while a similar increase in aggressiveness reduces it by 7 percentage points.[24] Finally, at the highest step (full provincial to deputy national or above), most trait coefficients fail to reach conventional levels of statistical significance, likely due to both limited sample size and the inherent uncertainty surrounding elite appointment decisions. However, we that perceived trustworthiness remains the trait with the most positive and precisely estimated

[23] As a robustness check, we also estimate a pooled model with interaction terms that allows the effects of facial traits to differ across promotion steps. The full results are reported in Table C.7 and Table C.8.

[24] By contrast, at the previous step (prefecture to deputy provincial), a one-standard-deviation increase in trustworthiness only raises promotion probability by 0.2 percentage points and a one standard deviation in aggressiveness reduces it by 0.8 percentage points.

effect. Overall, these results are in line with our argument that at the higher rungs of the power hierarchy, concerns about potential disloyalty and power abuse dominate, and the selectorate are hence more likely to favor candidates who exhibit a non-threatening and trustworthy demeanor.

Table 2: Results by Promotion Steps

| | DV: Promoted to Higher Level (1 = yes) | | |
|---|---|---|---|
| | (1) Prefecture to deputy provincial | (2) Deputy provincial to full provincial | (3) Full provincial to deputy national or above |
| Perceived attractiveness | 0.009 | 0.035 | -0.035 |
| | (0.033) | (0.060) | (0.101) |
| Perceived competence | 0.083* | 0.113 | 0.180 |
| | (0.035) | (0.071) | (0.103) |
| Perceived trustworthiness | 0.012 | 0.170* | 0.205* |
| | (0.034) | (0.068) | (0.103) |
| Perceived aggressiveness | -0.037 | -0.326** | 0.042 |
| | (0.038) | (0.073) | (0.096) |
| Han ethnicity | -0.010 | -0.074 | 0.015 |
| | (0.032) | (0.049) | (0.086) |
| Female | 0.122** | -0.012 | 0.024 |
| | (0.032) | (0.041) | (0.067) |
| College degree | 0.184** | 0.193** | 0.048 |
| | (0.063) | (0.037) | (0.169) |
| Graduate degree | 0.244** | 0.360** | 0.082 |
| | (0.063) | (0.036) | (0.181) |
| Birth year FE | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ |
| R-squared | 0.15 | 0.29 | 0.35 |
| Observations | 3716 | 1500 | 609 |

***Note***: This table presents the estimated effects of facial traits at each promotion step. Each column corresponds to a specific step (prefecture to deputy provincial; deputy provincial to full provincial; full provincial to deputy national or above). The dependent variable is a binary indicator for whether an official serving at the lower rank was promoted to the next rank by 2022. FE = fixed effects.
Standard errors clustered at the birth city level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$

### Results by Gender

We also investigate whether there are gender-specific preferences for facial traits. Table 3 presents the results separately for female and male officials.[25] We see that the estimates for female and male subsamples are similar in the trustworthiness and competence dimensions—but diverge in attractiveness and aggressiveness. In terms of attractiveness, male officials appear to benefit modestly from having an attractive look, whereas female officials experience a minor career penalty. This pattern contrasts with findings from the electoral context, where attractiveness is generally more advantageous for female candidates than for male candidates (Chiao, Bowman, and Gill 2008; Poutvaara, Jordahl, and Berggren 2009). However, it is consistent with anecdotal accounts that senior government leaders sometimes deliberately refrain from promoting good-looking female candidates to avoid speculative rumors of favoritism or inappropriate relationships (Chu 2022, 161–164). In contrast to attractiveness, an aggressive appearance seems to hurt the careers of male officials but benefits female officials. This may reflect the long-standing stereotype within the system that femininity (often associated with lower aggressiveness) correlates negatively with capabilities. Additionally, in a system where the most powerful positions are held by men, female subordinates may be viewed as less threatening to power holders compared to their male counterparts. This makes women's aggressive appearance more tolerable than men's.

## Robustness

We conducted a series of additional tests to evaluate the robustness of our findings. To ensure that our results are not driven by the specific method we used to construct the facial rating variables, we re-ran the main regressions using two alternative approaches: (1) a discrete version of the rating variables based on value quartiles (Figure B.1) and (2) rating variables based on the median rather than the mean of individual photo ratings (Table C.1). The main results remain largely unchanged. A potential alternative mechanism of face-based selection is that top decision-makers (or

[25] Gender-specific interaction estimates are reported in Appendix Table C.9. and Table C.10.

Table 3: Results by Candidates' Gender

| | Promotion | | Purge | |
|---|---|---|---|---|
| | (1) Female | (2) Male | (3) Female | (4) Male |
| Perceived attractiveness | -0.755 | 0.062 | -0.214 | -0.019 |
| | (0.524) | (0.064) | (0.146) | (0.029) |
| Perceived competence | 0.945 | 0.272** | -0.218 | -0.060* |
| | (0.516) | (0.069) | (0.154) | (0.025) |
| Perceived trustworthiness | 0.854 | 0.229** | -0.317* | -0.054 |
| | (0.614) | (0.074) | (0.139) | (0.030) |
| Perceived aggressiveness | 1.521* | -0.268** | 0.084 | -0.000 |
| | (0.697) | (0.073) | (0.198) | (0.028) |
| Han ethnicity | -0.320 | -0.030 | -0.062 | -0.010 |
| | (0.328) | (0.055) | (0.068) | (0.023) |
| College degree | -2.959** | 0.618** | -0.559** | 0.122** |
| | (0.594) | (0.154) | (0.202) | (0.029) |
| Graduate degree | -3.303** | 0.776** | -0.545** | 0.143** |
| | (0.603) | (0.160) | (0.163) | (0.028) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.83 | 0.26 | 0.77 | 0.13 |
| Observations | 249 | 3822 | 249 | 3822 |

***Note***: This table presents the estimated effects of facial traits separately for female and male officials. The specification is otherwise identical to Table 1. FE = fixed effects. Standard errors clustered at the birth city level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$

their underlings) might prefer candidates who resemble themselves, as perceived self-resemblance has been linked to trust and cooperation (Platek, Krill, and Wilson 2009). To address this possibility, we included three variables measuring the facial similarity between each official and three successive top leaders (Jiang Zemin, Hu Jintao, and Xi Jinping). Inclusion of these similarity measures does not substantively alter the main results (Table C.2).

As discussed earlier, reverse causality is a major alternative explanation for our findings. Higher-ranking officials might have better publicity staff who could produce photos depicting them as warm, competent, and non-aggressive. We address this issue in several ways. First, we constructed direct measures for image quality (face quality and photo blurriness) using Face++, a popular computer vision API (https://www.faceplusplus.com/), and included them as con-

trols in the regressions (Table C.3). The results remained robust to the inclusion of these controls. Second, we generated separate facial ratings from the two types of photos—official portraits and photos taken during meetings and public events. Official portraits are typically more heavily edited than meeting photos. If systematic editing efforts were directed toward cultivating specific visual impressions, we would expect significant differences in results between the two photo sources. Contrary to this expectation, however, we find largely comparable results (Table C.4). Third, we analyzed multiple photos of the same official taken at different stages of their careers to assess whether their observed facial traits changed after major promotions. We focused on non-military Politburo members of the 19th and 20th Central Committee, whose long careers provided sufficient photos from both current and past postings (an average of 10 photos per official). Our analysis suggests that ratings across all four dimensions remained largely stable throughout an official's career, with minimal changes following a promotion to the Politburo (Figure B.2 and Table C.5). These various tests increase our confidence that our main findings are unlikely to be solely driven by more intense image manipulation from officials who are already in higher-level offices.

## Relative Explanatory Power of Facial Appearances

Results from the preceding analysis provide strong evidence that facial appearances systematically influence the promotions and demotions of mid- and high-level officials in the Chinese government. An important question that remains, however, is how much explanatory power these appearance-based variables have relative to more commonly studied metrics, such as education, performance, and patron-client connections (Jia, Kudamatsu, and Seim 2015; Landry, Lü, and Duan 2018; Manion 2023; Shih 2008). To address this, we estimate Random Forest (RF) models that incorporate both facial ratings and a range of variables related to an official's performance, political connections, and demographic background. For performance, we calculate the average growth rates in GDP and fiscal revenue for each official during their tenure as city leaders.[26] For

[26]For officials who have not held city-level positions, no standardized method exists to assess their performance, so they are excluded from this analysis.

political connection, we measure the number of years an official served as a city leader under provincial leaders who later became Politburo members. Other background variables include an official's gender, education, ethnicity, and the latitude and longitude of their hometown.

In the RF model, each variable is assigned a variable importance (VI) score, which quantifies the reduction in predictive accuracy if the data for that variable is permuted while all others remain unchanged. Figure 4 displays the VI score for the main variables in descending order. For promotion outcomes, the four facial rating variables rank the highest among all variables in terms of VI scores. For purge outcome, the VI scores for facial ratings are somewhat lower than those for economic and fiscal performance,[27] but comparable to hometown locations and birth year, and noticeably higher than political connection, gender, education, and ethnicity.[28] While caution is warranted in interpreting the relative magnitude of VI scores—given the inherent challenges in measuring performance and connections—the general patterns revealed by the RF models are nonetheless informative. The results suggest that appearance-based assessments are not peripheral considerations but play a role in promotion decisions comparable to the traditionally recognized meritocratic and relational criteria.

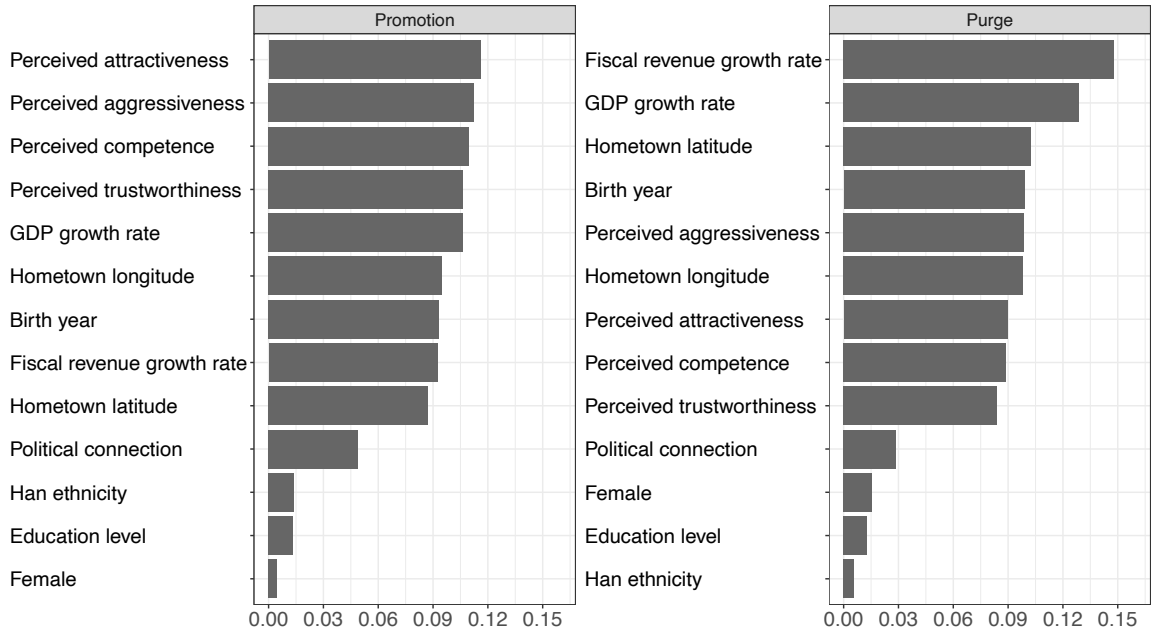
# Evidence from Conjoint Experiments with Simulated Photos

The observational data analysis presented thus far provides compelling evidence on the existence of strong but latent appearance-based selection preferences within the Chinese bureaucracy. However, two important gaps remain. First, since the deep learning models were trained on data

[27] Note that the high VI scores for fiscal revenue and GDP growth in the purge model may not suggest a performance-based story, as the estimated associations between these performance variables and purge are positive in a linear model. A more plausible interpretation is that officials serving in rapidly growing cities are more likely to become targets of corruption investigations.

[28] Unlike the baseline specification, we used a binary outcome variable for promotion in the RF model. The variable is coded as 1 if a city leader experienced a rank change and 0 otherwise. This approach avoids the severe class imbalance problem in multi-class rank models, which arises from the limited number of cases at higher ranks and the small overall sample size. Although the binary approach may also face a class imbalance issue, we address this using SMOTE (Synthetic Minority Oversampling Technique) to generate synthetic samples for the minority class. The binary model achieves strong predictive performance, with an accuracy of 86.5% and a weighted kappa of 0.729, both indicating high predictive power. For the purge model, a binary classification approach is naturally appropriate and yields a predictive accuracy of 88.0%, and a weighted kappa of 0.759

Figure 4: Relative Importance of Facial Appearances in Predicting Promotion and Purge




***Note*:** This figure displays the variable importance plot from Random Forests models. The horizontal bars indicate the importance score for each variable.

annotated by ordinary citizens, it is possible that the general public interprets officials' facial appearances differently than government insiders. Second, and more importantly, to the extent that individuals with different facial traits may differ in their actual personality or competence, the observed selection patterns may reflect substantive differences in real personal traits, rather than the influence of superficial impressions.

To address these limitations and verify that the posited selection preferences do exist at the individual level, we conducted a series of conjoint survey experiments with real government officials (n = 159).[29] In these experiments, we asked our subjects to review the profiles of two

[29]The design and protocol of the experimental study were approved by the Institutional Review Board at [ANONYMIZED INSTITUTION] (IRB-AAAV5254). Participants were recruited from subscribers to a popular social media account that provides proprietary career consultation services to civil servants and state employees. We distributed the survey link within the online subscriber chat groups managed by the account administrator. Participation was fully voluntary, though a small amount of monetary incentive was offered. Of the 179 initial respondents, we excluded those not employed in the state sector and those who failed to correctly answer two basic civil servant knowledge questions correctly, resulting in a sample with 159 valid observations. As shown in Figure B.1, despite being a convenience sample, the regional distribution of respondents aligns closely with the relative sizes of civil servant

hypothetical candidates and choose one for promotion. Each candidate profile includes a photo and a short description of the person's background. For the photos, we used synthetic frontal headshots generated by StyleGAN2, a state-of-the-art algorithm developed by NVIDIA to produce high-quality simulated images (Karras et al. 2020). StyleGAN2 models were trained separately for the highest-rated and the lowest-rated 3,000 photos in each of the four trait dimensions, producing eight distinct pools of synthetic images.[30] Drawing from these image pools, we created matched photo pairs in which the two photos differ mainly in one dimension but have very similar ratings in the other three.[31] In addition to a synthetic photo, we also included in each profile three attributes most commonly considered in promotion deliberations—birth year (1973, 1975, 1977, and 1979), educational attainment (bachelor's degree, master's degree, Ph.D degree), and work experience (state-owned enterprises, grassroots government, central government, or as a personal aide to senior leaders). Each subject evaluated eight pairs of candidate profiles (two pairs per dimension) where the photos and personal attributes were randomly combined and the order of appearance for pairs was also randomized. The interface and detailed wording of the experiment are provided in Appendix D.

Figure 5 reports the Average Marginal Component Effect (AMCE) of all attributes. The estimates (indicated by circles) represent the additional influence of each attribute on the probability (relative to the baseline group) that a profile would be selected for promotion, and the horizontal bars represent the 95% confidence intervals. The results from the conjoint experiment are highly consistent with findings from the observational analyses. Within paired profiles, candidates with higher-rated synthetic photos in perceived trustworthiness and competence, or lower perceived aggressiveness, are significantly more likely to be selected for promotion. Perceived attractiveness,

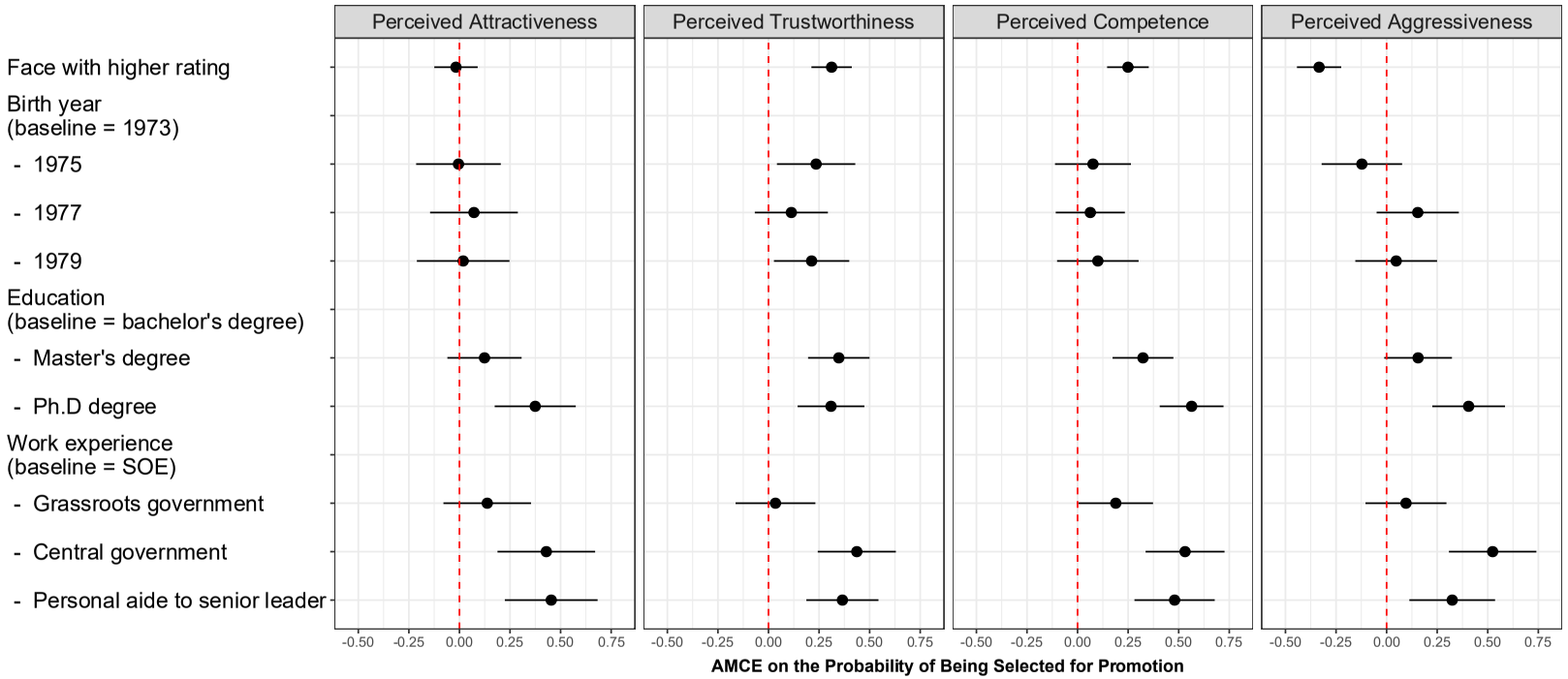
corps across provinces. For a similar conjoint design that examines the recruitment preferences for entry-level civil servants, see Liu (2018).

[30]Due to the limited number of female official photos in the dataset (1,582, which is below the minimum required for StyleGAN2 training), the experimental section focuses exclusively on male officials.

[31]To create matched photo pairs, we began by selecting the top 100 images in the high-rating pool of a target dimension. For each image, we calculated the sum of absolute differences in the other three dimensions and identified the 100 most similar matches from the low-rating pool of the same dimension. From these, we selected the pair with the largest difference in the target dimension while minimizing the differences in the control dimensions. In the end, two best pairs were selected for each dimension.

however, does not significantly impact promotion choices when the other three dimensions are controlled for. In terms of substantive magnitude, the premiums that our government respondents placed on more trustworthy and less aggressive appearances are substantial. These traits have an influence comparable to having a Ph.D (relative to a bachelor's degree) or the difference between grassroots government experience and serving as a personal aide to a senior leader. The benefit of appearing more competent, though somewhat smaller than trustworthiness and non-aggressiveness, also amounts to about 70% of the effect of a master's degree (relative to a bachelor's), or 47% of the effect of central government experience (relative to grassroots government). Collectively, these findings provide confirmatory individual-level evidence that perceptions of facial appearances exert a measurable and independent influence on promotion decisions—comparable to the effect of advanced degrees or favorable work experiences.

Figure 5: Conjoint Experiment Results on Individual Selection Preferences




***Note*:** This figure presents the results from four conjoint experiments on the promotion preferences of real government officials. In each experiment, respondents were shown photos and brief biographies of a pair of hypothetical officials and asked to choose one for promotion. The photos were drawn from pools of simulated images and matched on all dimensions except the one under comparison. The circles mark the point estimates for the effect of a given attribute on promotion choice and the horizontal bars indicate the 95% confidence intervals. The full numerical results are reported in Table D.1. SOE = state-owned enterprises.

# Conclusion

"The most interesting and astounding contradiction in life", as noted by Chester Barnard (1936), a pioneering scholar in the study of bureaucratic organizations, is the "constant insistence by nearly all people upon 'logic'...[and] 'sound reasoning' on the one hand, and on the other their inability to display it." This tension between modern regimes' desire for rationalization and the difficulty of realizing it in practice is particularly evident in the contemporary Chinese party-state. As our analysis demonstrates, despite long-standing efforts to institutionalize the selection process through rules, procedures, and systems, these efforts have not fully excluded the influence of primordial instincts and superficial impressions from personnel decisions. By examining the relationship between automated facial ratings and career outcomes for over 4,000 mid- and senior-level officials, we provide evidence that officials whose facial features convey impressions of competence, trustworthiness, and non-aggressiveness are significantly more likely to be promoted and face a lower risk of demotion than their peers. Perceived trustworthiness and non-aggressiveness are especially crucial for male officials and for promotions to the higher echelons of power. Furthermore, using Random Forest models and conjoint experiments, we demonstrate that facial appearance influences career outcomes to a degree comparable to conventional and "objective" metrics, such as educational attainment, economic performance, and political connections.

Our findings raise complicated normative implications. One major implication is about fairness. The presence of appearance-based selection naturally raises concerns that it may favor undeserving individuals and reinforce stereotypes and biases. However, an alternative argument could be that, since appearance is a universal marker possessed by a large number of individuals, it can sometimes be used to counteract the influence of other more exclusive traits, such as connections, and potentially help broaden the pool of candidates in the selection process. A second major issue is how appearance-based judgment affects the efficacy of selection. Here, a key question is how accurately real traits and character can be inferred from appearances. On the one hand, evolutionary psychologists argue that physical features may have historically served as evolutionary cues to ancestrally relevant traits, such as prowess, strength, and resilience—attributes potentially

valuable for certain political or institutional roles (Vugt and Grabo 2015). Some empirical studies suggest that facial appearances can indeed predict real-world behavioral tendencies in specific domains, such as aggressiveness (Třebický et al. 2013) and extroversion (Borkenau et al. 2009). On the other hand, some researchers have pointed out that face-based judgments often fail to surpass those made using more explicit heuristics, and that reliance on appearance may even hinder the use of other relevant information in making evaluations (Olivola and Todorov 2010b). This raises the critical question of whether the use of facial cues improves or undermines the quality of selection.[32] Of course, given the inherently uncertain nature of politics, we may never fully observe the "true" quality of a candidate until a moment of real challenge. After all, past qualifications and credentials might simply be another "face"—a professional self-presentation designed to create favorable impressions for a targeted audience. As demonstrated by numerous historical and contemporary cases—from Franklin Roosevelt during the Great Depression to Volodymyr Zelensky in the Russia-Ukraine War,[33] experiences in "normal politics" do not always predict leadership directions or effectiveness in extraordinary circumstances.

The finding that one of the powerful political bureaucracies on earth favors candidates with a trustworthy and non-threatening appearance invites an interesting comparison with the electoral selection process, where voters tend to prioritize traits such as attractiveness and competence rather than social warmth (Joo, Steen, and Zhu 2015; Todorov et al. 2005).[34] This discrepancy in preferences highlights important distinctions between electoral and bureaucratic systems. Modern elections are, at their core, pacified political warfare between coalitions of parties and interest groups (Laustsen and Petersen 2018; Przeworski 2018). Intense inter-group competition usually creates

[32] In an additional analysis, we show that local officials' facial traits are not significantly correlated with their economic or fiscal performance, suggesting that the substantive difference in performance between officials with different appearance may be limited (Table C.6).

[33] In the case of Roosevelt, despite being dismissed as a wealthy aristocrat who knew little of common suffering, he emerged as a champion for the "forgotten man" during the Great Depression, implementing some of the most radical redistribution programs in the US history (Brands 2009). More recently, as the comedian-turned-president of Ukraine, Zelensky was regarded as "completely out of depth" when taking office, but demonstrated highly effective leadership in the country's resistance against the invasion of the Russian military. See "Volodomyr Zelenskyy: From 'completely out of his depth' to Ukraine's president in war", Sky News, March 15, 2023, https://shorturl.at/xEr50.

[34] An notable exception is Chen, Jing, and Lee (2014), who find a positive interactive effect between competence and trustworthiness using candidates from 2008 U.S. Senate elections. However, their study exclusively involved undergraduate students, and their results did not show a significant main effect for trustworthiness.

a demand for leaders who appear dominant and assertive, as such individuals are perceived to be capable fighters who can better defend the interests of the ingroup (Spisak et al. 2012). In contrast, bureaucracies are hierarchical organizations designed to coordinate large-scale collective actions. They require members to have strong internal solidarity and adhere to formal rules and authority (Evans 1995). In this context, candidates who appear humble and dependable are often favored, as they are seen as less likely to disrupt organizational cohesion through disobedience or abuse of power.[35] This interpretation also aligns with findings from studies showing that trustworthiness as a facial trait is also valued in the selection processes of corporate hierarchies (Linke, Saribay, and Kleisner 2016) and in societies with more hierarchical cultural norms (Rule et al. 2010). The preferences for candidate appearances may thus reflect the fundamental differences in how power is organized and distributed across political systems.

The fact that manual ratings by ordinary citizens can predict high-level political selection offers an interesting perspective on mass-elite linkages in the absence of elections. It is well established that the Chinese government enjoys an unusually high level of trust among the citizens, but researchers continue to debate the sources of such trust (Tang 2016). While some scholars attribute it to political desirability bias (Nicholson and Huang 2022), others highlight the government's effort to deliver strong performances as a way of sustaining the regime's popularity (Yang and Zhao 2014). Our results, however, suggest another possible mechanism via public perceptions: Part of the public's trust in government may stem from the system's tendency to select trustworthy-looking individuals for high-level offices. Although public image may not be the most salient consideration in the selection of high-ranking officials, to the extent that human beings share similar impressions of facial traits, individuals perceived as trustworthy and non-aggressive by elite decision-makers will likely make similar impressions on ordinary citizens. This linkage may unintentionally help

[35] A parallel observation was made by Sahlins (1963) in his classic study of Melanesia "big men" and Polynesian chiefs. According to Sahlins, a decentralized, politically unintegrated tribal society like Melanesia's gave rise to leaders who relied on personal skills, charisma, and competitiveness to attract followers. In contrast, Polynesia's pyramidal society fostered a form of leadership characterized by humble, collegial chiefs whose authority "resided in the office...not made by the demonstration of personal superiority" (295). Sahlins further makes an incisive observation that the big men were "thoroughly bourgeois, so reminiscent of the free enterprising rugged individual of our own heritage" (289), whereas Polynesian chiefs bore a closer resemblance to the bureaucratic model of governance characteristic of communist states.

the regime project an image of benevolence and humility to the public. A similar mechanism might also help explain the cross-national pattern in which civil services typically often enjoy higher levels of public trust than legislatures, despite the latter being elected bodies (e.g., OECD 2024).

Methodologically, the deep learning approach adopted in this study enables us to conduct multi-dimensional facial trait assessment for a large number of political actors. This opens several promising avenues for future research. Within a single national polity, similar methods could be used to explore how preferences for appearance vary across time, regions, and functional roles. For example, do different sectors of the state (e.g., police vs. civilian administration) prefer officials with different facial profiles? How do preferences for facial appearances evolve over time as a political organization progresses through its life cycles (Downs 1967) or distinct stages of institutionalization (Huntington 1968)? Researchers may also use facial traits as explanatory variables to investigate their potential impact on an official's substantive policy preferences or administrative performance. Beyond single-country studies, an even more exciting direction is to leverage faces as a universal medium to construct measures for a global sample of political figures, such as their perceived personality traits, interpersonal dynamics, or emotional states during key political events. Quantifying and comparing these nuanced yet important cues cross-nationally could allow researchers to look beyond the rational facade of formal institutions and see a deeper layer of politics—one rooted in feelings, instincts, and perceptions.

# References


Albright, Linda, et al. 1997. "Cross-cultural Consensus in Personality Judgments." *Journal of Personality and Social Psychology* 72 (3).

Alfertshofer, Michael, et al. 2024. "The "Central Tendency Bias" in the assessment of facial attractiveness in group-based and individual ratings—A survey-based study in 727 volunteers." *Journal of Plastic, Reconstructive & Aesthetic Surgery* 92:264–275.

Antonakis, John, and Olaf Dalgas. 2009. "Predicting Elections: Child's Play!" *Science* 323 (5918): 1183–1183.

Banducci, Susan A., et al. 2008. "Ballot Photographs as Cues in Low-Information Elections." *Political Psychology* 29, no. 6 (): 903–917.

Bar, Moshe, Maital Neta, and Heather Linz. 2006. "Very first impressions." *Emotion* 6 (2): 269–278.

Barnard, Chester I. 1936. *Mind in Everyday Affairs: An Examination Into Logical and Non-logical Thought Processes*. United States: Guild of Brackett Lecturers.

— . 1938. *The Functions of the Executive*. 1st. Cambridge, MA: Harvard University Press.

Bell, D.A. 2016. *The China Model: Political Meritocracy and the Limits of Democracy*. Princeton University Press.

Benjamin, Daniel J, and Jesse M Shapiro. 2009. "Thin-Slice Forecasts of Gubernatorial Elections." *Review of Economics and Statistics* 91 (3): 523–536.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara. 2010. "The looks of a winner: Beauty and electoral success." *Journal of Public Economics* 94, no. 1–2 (): 8–15.

Besley, Timothy. 2005. "Political Selection." *Journal of Economic Perspectives* 19, no. 3 (): 43–60.

Borkenau, Peter, et al. 2009. "Extraversion is accurately perceived after a 50-ms exposure to a face." *Journal of Research in Personality* 43, no. 4 (): 703–706.

Brands, H. W. 2009. *Traitor to His Class: The Privileged Life and Radical Presidency of Franklin Delano Roosevelt*. Vintage.

Brandt, Anthony. 1980. "Face Reading: The Persistence of Physiognomy." *Psychologist Today* 90:90–96.

Budesheim, Thomas Lee, and Stephen J. DePaola. 1994. "Beauty or the Beast? The Effects of Appearance, Personality, and Issue Information on Evaluations of Political Candidates." *Personality and Social Psychology Bulletin* 20, no. 4 (): 339–348.

Castelli, Luigi, et al. 2009. "The effects of perceived competence and sociability on electoral outcomes." *Journal of Experimental Social Psychology* 45 (5): 1152–1155.

Chen, Fang Fang, Yiming Jing, and Jeong Min Lee. 2014. "The looks of a leader: Competent and trustworthy, but not dominant." *Journal of Experimental Social Psychology* 51:27–33.

Chen, Yun. 1995. *Selected Works of Chen Yun* 陈云文选. Vol. 3. Beijing, China: People's Publishing House 人民出版社.

Chiao, Joan Y., Nicholas E. Bowman, and Harleen Gill. 2008. "The Political Gender Gap: Gender Bias in Facial Inferences that Predict Voting Behavior." Ed. by Laurie Santos. *PLoS ONE* 3 (10): e3666.

Chu, Chaoxin. 2022. *I Actually Still Want to Improve* 我，其实还想进步. Beijing, China: People's Daily Press 人民日报出版社.

Dal Bó, Ernesto, and Frederico Finan. 2018. "Progress and Perspectives in the Study of Political Selection." *Annual Review of Economics* 10 (1): 541–575.

Ding, Iza Yue. 2022. *The Performative State: Public Scrutiny and Environmental Governance in China*. Cornell University Press.

Dotsch, Ron, et al. 2008. "Ethnic Out-Group Faces Are Biased in the Prejudiced Mind." *Psychological Science* 19 (10): 978–980.

Downs, Anthony. 1967. *Inside Bureaucracy*. Boston: Little Brown / Company.

Duchaine, Brad, and Galit Yovel. 2015. "A Revised Neural Framework for Face Processing." *Annual Review of Vision Science* 1 (1): 393–416.

Egorov, Georgy, and Konstantin Sonin. 2011. "Dictators and Their Viziers: Endogenizing the Loyalty-competence Trade-off." *Journal of the European Economic Association* 9 (5): 903–930.

Eisenbruch, Adar B., et al. 2024. "US adults accurately assess Hadza and Tsimane men's hunting ability from a single face photograph." *Evolution and Human Behavior* 45 (4): 106598.

Epstein, Seymour. 1994. "Integration of the Cognitive and the Psychodynamic Unconscious." *American psychologist* 49 (8): 709.

Evans, Peter B. 1995. *Embedded Autonomy: States and Industrial Transformation*. Book collections on Project MUSE. Princeton University Press.

Fan, Bonai, et al. 2018. "Rational Civil Servant Interviewers: Evidence from an Event-Related Potential Study of Beauty Premiums in Chinese Civil Servant Interviews." *Experimental Brain Research* 236, no. 12 (): 3159–3168.

Fiske, Susan T., Amy J.C. Cuddy, and Peter Glick. 2007. "Universal dimensions of social cognition: warmth and competence." *Trends in Cognitive Sciences* 11 (2): 77–83.

Freeman, J. B., et al. 2009. "The Neural Basis of Categorical Face Perception: Graded Representations of Face Gender in Fusiform and Orbitofrontal Cortices." *Cerebral Cortex* 20 (6): 1314–1322.

Fukuyama, Francis. 2014. *Political Order and Political Decay: From the Industrial Revolution to the Globalization of Democracy*. Farrar, Straus / Giroux.

Gallagher, Michael, and Michael Marsh, eds. 1988. *Candidate Selection in Comparative Perspective*. London: Sage Publications.

Giacomin, Miranda, and Nicholas O. Rule. 2020. "How Static Facial Cues Relate to Real-world Leaders' Success: A Review and Meta-analysis." *European Review of Social Psychology* 31, no. 1 (): 120–148.

Gigerenzer, Gerd. 2007. *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking Penguin.

Grindle, Merilee S. 2012. *Jobs for the Boys: Patronage and the State in Comparative Perspective*. Cambridge, MA: Harvard University Press.

Hassin, Ran, and Yaacov Trope. 2000. "Facing faces: Studies on the cognitive aspects of physiognomy." *Journal of Personality and Social Psychology* 78 (5): 837–852.

Haxby, James V., Elizabeth A. Hoffman, and M.Ida Gobbini. 2000. "The distributed human neural system for face perception." *Trends in Cognitive Sciences* 4 (6): 223–233.

Hirao, Naoyasu, et al. 2021. "Reliability of online surveys in investigating perceptions and impressions of faces." *Frontiers in Psychology* 12:733405.

Huang, Jing. 2000. *Factionalism in Chinese Communist Politics*. Cambridge University Press.

Huntington, Samuel P. 1968. *Political Order in Changing Societies*. New Haven, CT: Yale University Press.

Jack, Rachael E., and Philippe G. Schyns. 2015. "The Human Face as a Dynamic Tool for Social Communication." *Current Biology* 25 (14): R621–R634.

Janvry, Alain de, et al. 2023. "Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China." *American Economic Review* 113 (3): 766–799.

Jeong, Dahyeon, et al. 2023. "Exhaustive or exhausting? Evidence on respondent fatigue in long surveys." *Journal of Development Economics* 161:102992.

Jia, Ruixue, Masayuki Kudamatsu, and David Seim. 2015. "Political Selection in China: The Complementary Roles of Connections and Performance." *Journal of the European Economic Association* 13 (4): 631–668.

Jiang, Junyan, and Zhaotian Luo. 2021. "Leadership Styles and Political Survival of Chinese Communist Party Elites." *The Journal of Politics* 83 (2): 777–782.

Joo, Jungseock, Francis F. Steen, and Song-Chun Zhu. 2015. "Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face." In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 3712–3720. IEEE, IEEE.

Judge, Timothy A., et al. 2002. "Personality and leadership: A qualitative and quantitative review." *Journal of Applied Psychology* 87 (4): 765–780.

Kahneman, Daniel, and Gary Klein. 2009. "Conditions for intuitive expertise: A failure to disagree." *American Psychologist* 64 (6): 515–526.

Karras, Tero, et al. 2020. *Analyzing and Improving the Image Quality of StyleGAN*.

Keller, Franziska Barbara. 2016. "Moving beyond Factions: Using Social Network Analysis to Uncover Patronage Networks among Chinese Elites." *Journal of East Asian Studies* 16 (1): 17–41.

Key, V. O. 1956. *American State Politics: An Introduction*. New York: Alfred A. Knopf.

Khaneman, Daniel. 2003. "A Perspective on Judgement and Choice." *American Psychologist* 58:697–720.

Landry, Pierre F., Xiaobo Lü, and Haiyan Duan. 2018. "Does Performance Matter? Evaluating Political Selection Along the Chinese Administrative Ladder." *Comparative Political Studies* 51, no. 8 (): 1074–1105.

Laustsen, Lasse, and Michael Bang Petersen. 2015. "Does a competent leader make a good friend? Conflict, ideology and the psychologies of friendship and followership." *Evolution and Human Behavior* 36 (4): 286–293.

— . 2018. "When the Party Decides: The Effects of Facial Competence and Dominance on Internal Nominations of Political Candidates." *Evolutionary Psychology* 16 (2).

Lawson, Chappell, et al. 2010. "Looking like a Winner: Candidate Appearance and Electoral Success in New Democracies." *World Politics* 62 (4): 561–593. Visited on 08/21/2024.

LeCun, Yann, et al. 1989. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1 (4): 541–551.

Lenz, Gabriel S, and Chappell Lawson. 2011. "Looking the part: Television leads less informed citizens to vote based on candidates' appearance." *American Journal of Political Science* 55 (3): 574–589.

Li, Hongbin, and Li-An Zhou. 2005. "Political turnover and economic performance: the incentive role of personnel control in China." *Journal of Public Economics* 89 (9-10): 1743–1762.

Lin, Luojun, Lingyu Liang, and Lianwen Jin. 2019. "Regression Guided by Relative Ranking Using Convolutional Neural Network ($R^3$3 CNN) for Facial Beauty Prediction." *IEEE Transactions on Affective Computing* 13 (1): 122–134.

Ling, Leng, Danglun Luo, and Guoman She. 2019. "Judging a Book by Its Cover: The Influence of Physical Attractiveness on the Promotion of Regional Leaders." *Journal of Economic Behavior &amp Organization* 158:1–14.

Linke, Lenka, S. Adil Saribay, and Karel Kleisner. 2016. "Perceived trustworthiness is associated with position in a corporate hierarchy." *Personality and Individual Differences* 99 (): 22–27.

Little, Anthony C., et al. 2007. "Facial appearance affects voting decisions." *Evolution and Human Behavior* 28 (1): 18–27.

Liu, Hanzhang. 2023. "Meritocracy as Authoritarian Co-Optation: Political Selection and Upward Mobility in China." *American Political Science Review* 118 (4): 1856–1872.

— . 2018. "The Logic of Authoritarian Political Selection: Evidence from a Conjoint Experiment in China." *Political Science Research and Methods* 7 (04): 853–870.

Lord, Robert G., Christy L. de Vader, and George M. Alliger. 1986. "A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures." *Journal of Applied Psychology* 71 (3): 402–410.

Madison, James. 1788 [2008]. "Federalist No. 57: The Alleged Tendency of the New Plan to Elevate the Few at the Expense of the Many Considered." In *The Federalist Papers*, ed. by Alexander Hamilton, John Jay, and James Madison, 281–286. Reprinted 1961. New York: Oxford University Press.

Manion, Melanie. 2023. *Political Selection in China: Rethinking Foundations and Findings*. Elements in Politics and Society in East Asia. Cambridge University Press.

Mattes, Kyle, et al. 2010. "Predicting election outcomes from positive and negative trait assessments of candidate images." *Political Psychology* 31 (1): 41–58.

Mazur, Allan, Julie Mazur, and Caroline Keating. 1984. "Military Rank Attainment of a West Point Class: Effects of Cadets' Physical Features." *American Journal of Sociology* 90, no. 1 (): 125–150.

McGregor, Richard. 2010. *The Party: The Secret World of China's Communist Rulers*. New York: Harper Collins.

Merton, Robert. K. 1940. "Bureaucratic Structure and Personality." *Social Forces* 18, no. 4 (): 560–568.

Moe, Terry M. 2012. "Delegation, Control, and the Study of Public Bureaucracy." *The Forum* 10 (2).

Moore, David W., and B. Thomas Trout. 1978. "Military Advancement: The Visibility Theory of Promotion." *American Political Science Review* 72 (2): 452–468.

Mueller, Ulrich, and Allan Mazur. 1996. "Facial Dominance of West Point Cadets as a Predictor of Later Military Rank." *Social Forces* 74, no. 3 (): 823.

Nathan, Andrew J. 1973. "A Factionalism Model for CCP Politics." *The China Quarterly* 53 (): 34–66.

Nicholson, Stephen P., and Haifeng Huang. 2022. "Making the List: Reevaluating Political Trust and Social Desirability in China." *American Political Science Review* 117, no. 3 (): 1158–1165.

OECD. 2024. *OECD Survey on Drivers of Trust in Public Institutions – 2024 Results: Building Trust in a Complex Policy Environment*. Paris: OECD Publishing. https://doi.org/10.1787/9a20554b-en.

Olivola, Christopher Y., and Alexander Todorov. 2010a. "Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting." *Journal of Nonverbal Behavior* 34, no. 2 (): 83–110.

— . 2010b. "Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences." *Journal of Experimental Social Psychology* 46, no. 2 (): 315–324.

Oosterhof, Nikolaas N., and Alexander Todorov. 2008. "The functional basis of face evaluation." *Proceedings of the National Academy of Sciences* 105, no. 32 (): 11087–11092.

Platek, Steven M., Austen L. Krill, and Benjamin Wilson. 2009. "Implicit Trustworthiness Ratings of Self-Resembling Faces Activate Brain Centers Involved in Reward." *Neuropsychologia* 47, no. 1 (): 289–293.

Plato. 1968. *The Republic*. Trans. by Desmond Lee. Penguin Classics.

Polanyi, Michael. 1967. *The Tacit Dimension*. United Kingdom: Anchor Books.

Poutvaara, Panu, Henrik Jordahl, and Niclas Berggren. 2009. "Faces of politicians: Babyfacedness predicts inferred competence but not electoral success." *Journal of Experimental Social Psychology* 45 (5): 1132–1135.

Przeworski, Adam. 2018. *Why Bother With Elections?* Polity Press.

Rauch, James E, and Peter B Evans. 2000. "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries." *Journal of Public Economics* 75, no. 1 (): 49–71.

Rhodes, Gillian. 2006. "The Evolutionary Psychology of Facial Beauty." *Annual Review of Psychology* 57, no. 1 (): 199–226.

Rosenberg, Shawn W., et al. 1986. "The Image and the Vote: The Effect of Candidate Presentation on Voter Preference." *American Journal of Political Science* 30 (1): 108–127.

Rudolph, Lloyd I., and Susanne Hoeber Rudolph. 1979. "Authority and Power in Bureaucratic and Patrimonial Administration: A Revisionist Interpretation of Weber on Bureaucracy." *World Politics* 31 (2): 195–227.

Rule, Nicholas O., and Nalini Ambady. 2011. "Face and Fortune: Inferences of Personality from Managing Partners' Faces Predict Their Law Firms' Financial Success." *The Leadership Quarterly* 22, no. 4 (): 690–696.

— . 2008. “The Face of Success.” *Psychological Science* 19, no. 2 (): 109–111.

Rule, Nicholas O., et al. 2010. “Polling the face: Prediction and consensus across cultures.” *Journal of Personality and Social Psychology* 98 (1): 1–15.

Sahlins, Marshall D. 1963. “Poor Man, Rich Man, Big-Man, Chief: Political Types in Melanesia and Polynesia.” *Comparative Studies in Society and History* 5 (3): 285–303.

Selznick, Philip. 1952. *The Organizational Weapon: A Study of Bolshevik Strategy and Tactics*. RAND Corporation.

Shih, Victor Chung-Hon. 2008. ““Nauseating” Displays of Loyalty: Monitoring the Factional Bargain through Ideological Campaigns in China.” *The Journal of Politics* 70, no. 4 (): 1177–1192.

Shih, Victor, Christopher Adolph, and Mingxing Liu. 2012. “Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China.” *American Political Science Review* 106, no. 1 (): 166–187.

Sigelman, Lee, Carol K. Sigelman, and Christopher Fowler. 1987. “A Bird of a Different Feather? An Experimental Investigation of Physical Attractiveness and the Electability of Female Candidates.” *Social Psychology Quarterly* 50 (1): 32.

Simon, Herbert.A. 1947. *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*. Macmillan Company.

Spisak, Brian R., et al. 2012. “Warriors and Peacekeepers: Testing a Biosocial Implicit Leadership Hypothesis of Intergroup Relations Using Masculine and Feminine Faces.” *PLoS ONE* 7, no. 1 (): e30399.

Stanovich, Keith E, and Richard F West. 2000. “Advancing the Rationality Debate.” *Behavioral and brain sciences* 23 (5): 701–717.

Sutherland, Clare A. M., et al. 2015. “Personality judgments from everyday images of faces.” *Frontiers in Psychology* 6.

Sutherland, Clare A.M., et al. 2013. “Social inferences from faces: Ambient images generate a three-dimensional model.” *Cognition* 127 (1): 105–118.

Sy, Thomas. 2010. “What Do You Think of Followers? Examining the Content, Structure, and Consequences of Implicit Followership Theories.” *Organizational Behavior and Human Decision Processes* 113, no. 2 (): 73–84.

Tang, Wenfang. 2016. *Populist Authoritarianism: Chinese Political Culture and Regime Sustainability*. Oxford, UK: Oxford University Press.

Thiel, Sandra Van, and Frans L. Leeuw. 2002. “The Performance Paradox in the Public Sector.” *Public Performance & Management Review* 25 (3): 267–281. Visited on 08/30/2024.

Todorov, Alexander. 2017. *Face Value: The Irresistible Influence of First Impressions*. Princeton, NJ: Princeton University Press.

Todorov, Alexander, Manish Pakrashi, and Nikolaas N. Oosterhof. 2009. “Evaluating Faces on Trustworthiness After Minimal Time Exposure.” *Social Cognition* 27 (6): 813–833.

Todorov, Alexander, et al. 2005. "Inferences of Competence from Faces Predict Election Outcomes." *Science* 308, no. 5728 (): 1623–1626.

Todorov, Alexander, et al. 2015. "Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance." *Annual Review of Psychology* 66 (Volume 66, 2015): 519–545.

Třebický, Vít, et al. 2013. "Perceived Aggressiveness Predicts Fighting Performance in Mixed-Martial-Arts Fighters." *Psychological Science* 24 (9): 1664–1672.

Vugt, Mark Van, and Allen E. Grabo. 2015. "The Many Faces of Leadership: An Evolutionary-Psychology Approach." *Current Directions in Psychological Science* 24, no. 6 (): 484–489.

Wagner, Alexander F. 2011. "Loyalty and competence in public agencies." *Public Choice* 146 (1/2): 145–162. Visited on 08/23/2024.

Wang, Zhenyu M., Tao Li, and Rodrigo Praino. 2024. "Is there a prototype leader look? Evidence from the photos of Chinese local leaders." *The Leadership Quarterly* 35 (4): 101785.

Weber, Max. 1921 [1978]. *Economy and Society: An Outline of Interpretive Sociology*. Ed. by Guenther Roth and Claus Wittich. University of California Press.

Willis, Janine, and Alexander Todorov. 2006. "First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face." *Psychological Science* 17 (7): 592–598.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Wong, Stan Hok-wui, and Yu Zeng. 2017. "Do inferences of competence from faces predict political selection in authoritarian regimes? Evidence from China." *Social Science Research* 66 (): 248–263.

Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature* 49 (4): 1076–1151.

Yang, Hongxing, and Dingxin Zhao. 2014. "Performance Legitimacy, State Autonomy and China's Economic Miracle." *Journal of Contemporary China* 24, no. 91 (): 64–82.

Yao, Yang, and Muyang Zhang. 2015. "Subnational leaders and economic growth: evidence from Chinese cities." *Journal of Economic Growth* 20 (4): 405–436.

Zebrowitz, Leslie A. 1997. *Reading Faces: window to the soul?* 271. Westview Press.

Zebrowitz, Leslie A., and Joann M. Montepare. 2005. "Appearance DOES Matter." *Science* 308 (5728): 1565–1566.

Zebrowitz, Leslie A., Joann M. Montepare, and Hong Keun Lee. 1993. "They don't all look alike: Individuated impressions of other racial groups." *Journal of Personality and Social Psychology* 65 (1): 85–101.

Zeng, Qingjie. 2015. "Democratic Procedures in the CCP's Cadre Selection Process: Implementation and Consequences." *The China Quarterly* 225 (): 73–99.

# Online Appendix

## Table of Contents

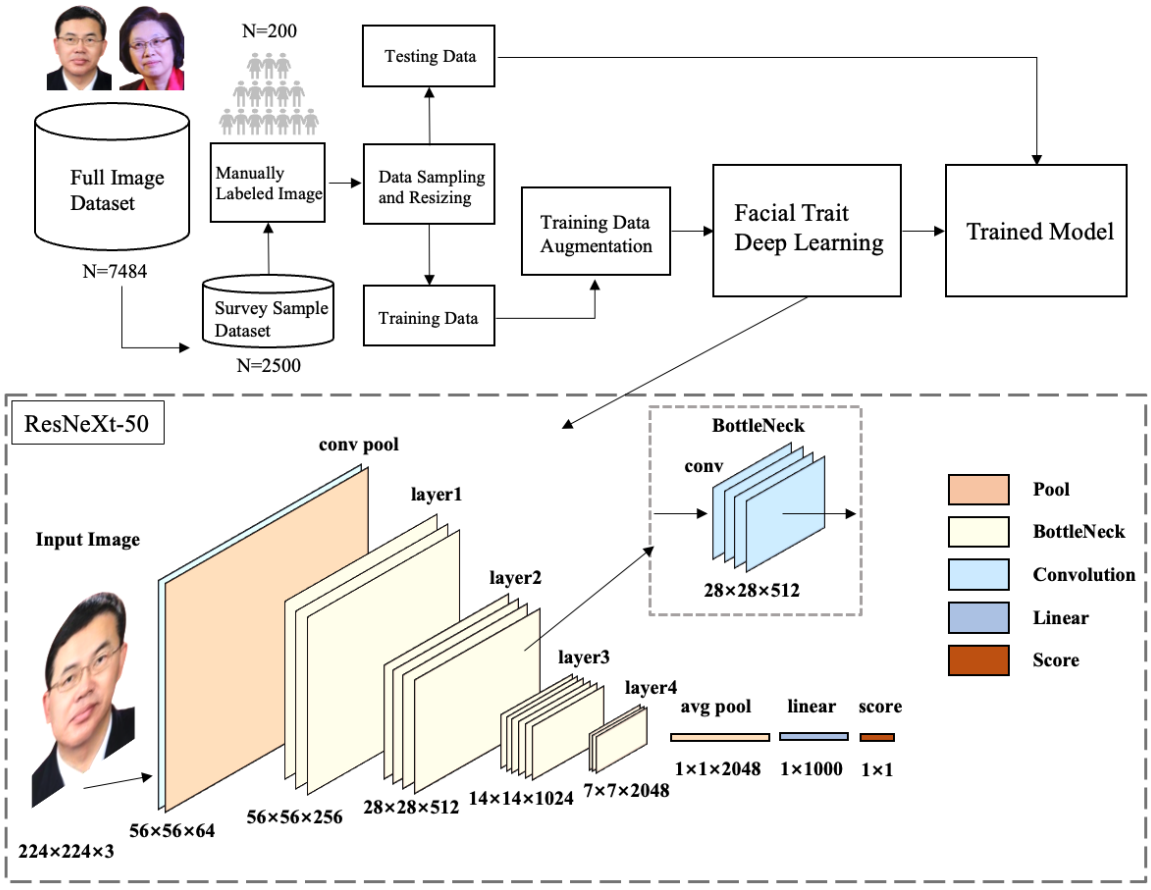
# A Summary Statistics

Table A.1: Summary Statistics

| Variable | N | Mean | S.D. | Min | Max |
|---|---|---|---|---|---|
| *Panel A: leaders' characteristics* | | | | | |
| Highest political rank | 4,174 | 1.41 | 0.83 | 1 | 5 |
| Demoted, arrested or received party disciplinary sanctions (1 = yes) | 4,174 | 0.09 | 0.29 | 0 | 1 |
| Female | 4,174 | 0.06 | 0.24 | 0 | 1 |
| Education level | 4,174 | 1.74 | 0.47 | 0 | 2 |
| Han ethnicity | 4,174 | 0.88 | 0.33 | 0 | 1 |
| City leaders: average GDP growth rate | 2,277 | -0.11 | 0.07 | -0.30 | 0.68 |
| City leaders: average fiscal income growth rate | 2,016 | 0.15 | 0.12 | -0.31 | 0.67 |
| City leaders: Connections with Politburo members | 2,279 | 1.13 | 2.38 | 0.00 | 16.01 |
| *Panel B: facial traits* | | | | | |
| Perceived attractiveness | 4,157 | 2.78 | 0.22 | 1.70 | 3.57 |
| Perceived competence | 4,157 | 3.05 | 0.22 | 1.74 | 3.91 |
| Perceived trustworthiness | 4,157 | 3.12 | 0.19 | 2.04 | 4.28 |
| Perceived aggressiveness | 4,157 | 2.78 | 0.16 | 1.82 | 3.62 |
| Forehead ratio | 4,154 | 0.25 | 0.01 | 0.20 | 0.28 |
| Eyebrow position | 4,154 | 0.13 | 0.01 | 0.08 | 0.17 |
| Eye size | 4,154 | 0.10 | 0.01 | 0.08 | 0.12 |
| Facial width-to-height ratio | 4,154 | 2.33 | 0.12 | 1.88 | 2.84 |
| Nose length | 4,154 | 0.27 | 0.01 | 0.24 | 0.31 |
| Nose width | 4,154 | 0.16 | 0.01 | 0.13 | 0.20 |
| Nose width-to-height ratio | 4,154 | 0.59 | 0.03 | 0.47 | 0.74 |
| Mouth width | 4,154 | 0.29 | 0.02 | 0.22 | 0.39 |
| Lip thickness | 4,154 | 0.08 | 0.01 | 0.07 | 0.11 |
| Chin length | 4,154 | 0.19 | 0.01 | 0.12 | 0.23 |
| Chin width | 4,154 | 0.38 | 0.02 | 0.29 | 0.43 |
| *Panel C: respondents' characteristics in conjoint* | | | | | |
| Female | 159 | 0.30 | 0.46 | 0 | 1 |
| Education level | 159 | 1.28 | 0.50 | 0 | 2 |
| Age group (1 = 20-30, 2 = 30-40, 3 = 40-50, 4= 50-60) | 159 | 1.78 | 0.46 | 1 | 4 |
| Employment type (1 = SOEs, 2 = Party and government organs) | 159 | 1.50 | 0.50 | 1 | 2 |
| Administrative rank (1 = none, 2 = deputy division and below, 3 = division level) | 159 | 1.52 | 0.51 | 1 | 3 |

# B Details on Automated Facial Rating Procedures and Output

Figure B.1: Schematic Representation of the Training Process.




## B.1 Information about Manual Rating Process

We first present raters with a brief prompt explaining the task and then give them 125 headshots to rate across four dimensions. Of these 125 photos, 25 are fixed and rated by all raters, while the remaining 100 are randomly selected from the pool of the 2,500 headshots excluding the 25 fixed ones. The 125 photos are displayed to raters, who evaluate them in one dimension at a time before moving on to the next dimension to rate the same set of photos again. The order of the photos is randomized for each dimension. The wording of the prompt and an example of a photo being rated are as follows.

**Prompt** “请观察以下照片中人物，根据您对他/她相貌的第一印象，从1至5进行打分” (“Please rate the person in the following photo on a scale from 1 to 5 based on your first impression of their appearance.”)

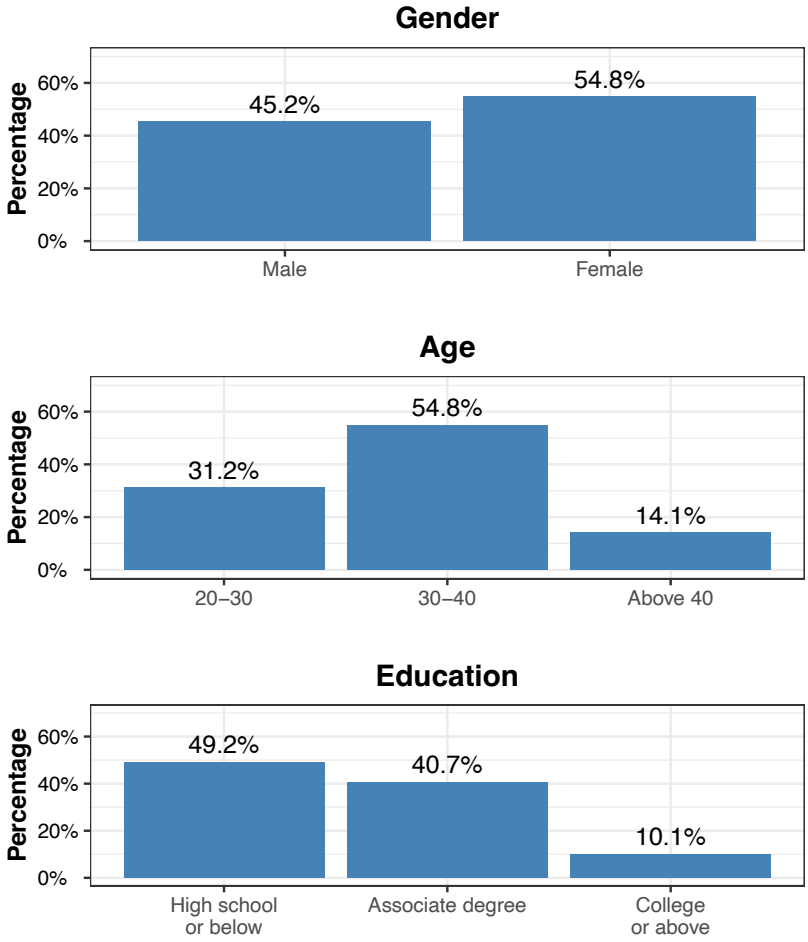
**Attractiveness question** 您认为图中人物有多好看？请您为他（她）的颜值打分（5代表非常好看，1代表非常不好看） (Does the person in the photo look attractive to you? Please rate his/her appearance on a scale from 1 to 5 (5 represents very attractive, 1 represents not attractive at all.))

**Competence question** 图中人物是否给您聪明精干的感觉？请从1到5打分（5代表非常聪明精干，1代表非常不聪明精干） (Does the person in the photo look competent to you? Please rate his/her appearance on a scale from 1 to 5 (5 represents very competent, 1 represents not competent at all.))

**Trustworthiness question** 图中人物是否给您忠厚可靠的感觉？请从1到5打分（5代表非常忠厚可靠，1代表非常不忠厚可靠） (Does the person in the photo look trustworthy to you? Please rate his/her appearance on a scale from 1 to 5 (5 represents very trustworthy, 1 represents not trustworthy at all.))

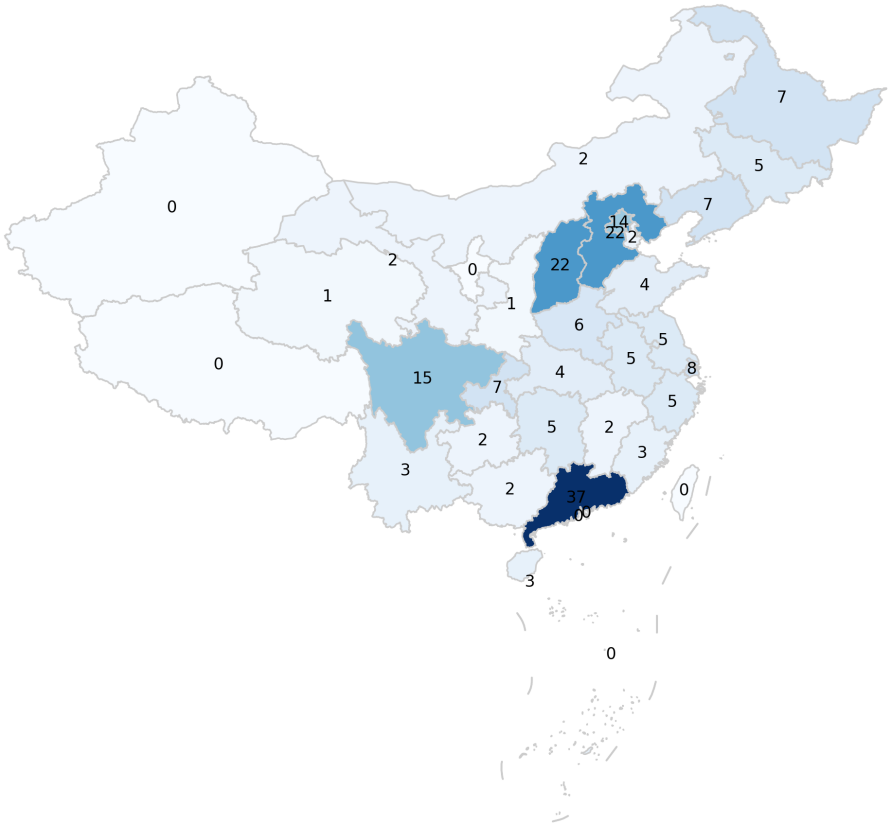
**Aggressiveness question** 图中人物是否给您凶狠强悍的感觉？请从1到5打分（5代表非常凶狠强悍，1代表非常不凶狠强悍） (Does the person in the photo look aggressive to you? Please rate his/her appearance on a scale from 1 to 5 (5 represents very aggressive, 1 represents not aggressive at all.))

Figure B.2: Demographic Background of Raters for the Training Dataset




***Note*:** This figure displays the distribution of key demographic attributes for human raters (n = 199).

Figure B.3: Distribution of Raters by Province




***Note*:** This figure displays the distribution of raters by their birth province (n = 199), with the number in each province representing the count of raters. The map shows a relatively balanced distribution of raters across China's northern, southern, eastern, and western regions.

Table B.1: Inter-rater Reliability

| Facial Dimension | Cronbach's $\alpha$ | Guttman's $\lambda_6$ |
|---|---|---|
| Perceived attractiveness | 0.91 | 0.92 |
| Perceived trustworthiness | 0.69 | 0.73 |
| Perceived competence | 0.71 | 0.74 |
| Perceived aggressiveness | 0.84 | 0.86 |

***Note***: This table provides information about the inter-rater reliability for each facial dimension. We use two measures: Cronbach's $\alpha$ and Guttman $\lambda_6$. Both range from 0 to 1, with larger values indicating greater agreement among raters. The values are calculated based on the 25 photos rated by all raters.

## B.2 Validation of Rater Familiarity with Officials

As noted in the main text, a key assumption of our human-rating task is that officials in our annotation sample are largely unrecognizable to the public, minimizing familiarity bias. To test this, we conducted a supplementary recognition survey, summarized below.

The survey used a pool of 227 officials' images covering all ranks from prefectural leaders to Politburo Standing Committee members, including a sample of female officials. We recruited 186 respondents online and applied post-stratification weights to match the provincial distribution of the original rater pool (see Table B.2). Each respondent viewed a random subset of 20 photos and answered two questions for each: whether they had seen the person before and whether they could identify them (name or position).

Respondents reported a general sense of familiarity with officials (50-55%), consistent across ranks (Figure B.4). However, actual recognition was extremely low (Figure B.5). For Ranks 1-3, the correct identification rate was **1.4%**, **0.7%**, and **1.5%**, respectively, with overlapping 95% confidence intervals. Even for Rank 4 and PSC members (Rank 5), recognition rates were only **2.3%** and **3.6%**. The contrast between the relatively high "seen" rate and near-zero correct identification indicates that respondents were reacting to general facial archetypes rather than specific individuals.

This survey confirms that officials from Ranks 1-3 are virtually anonymous to the public, validating their inclusion in the human-rating task. The facial trait scores in our main analysis therefore reflect first impressions of the images rather than prior knowledge or familiarity with the officials.

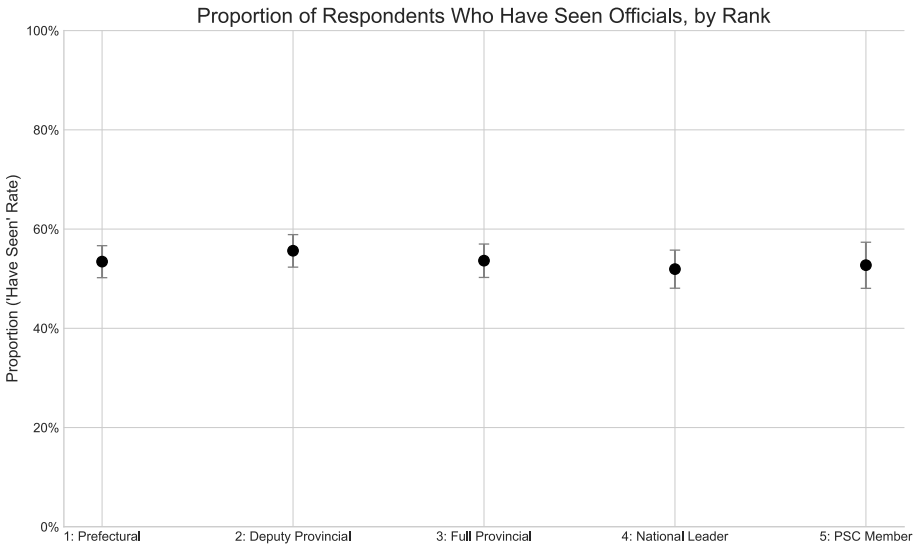



Figure B.4: Proportion of Respondents Reporting Having "Seen" Officials, by Rank

Notes: The figure displays the proportion of respondents who answered "Yes" to having seen the official before. Dots represent the point estimate, and the vertical bars represent 95% confidence intervals.

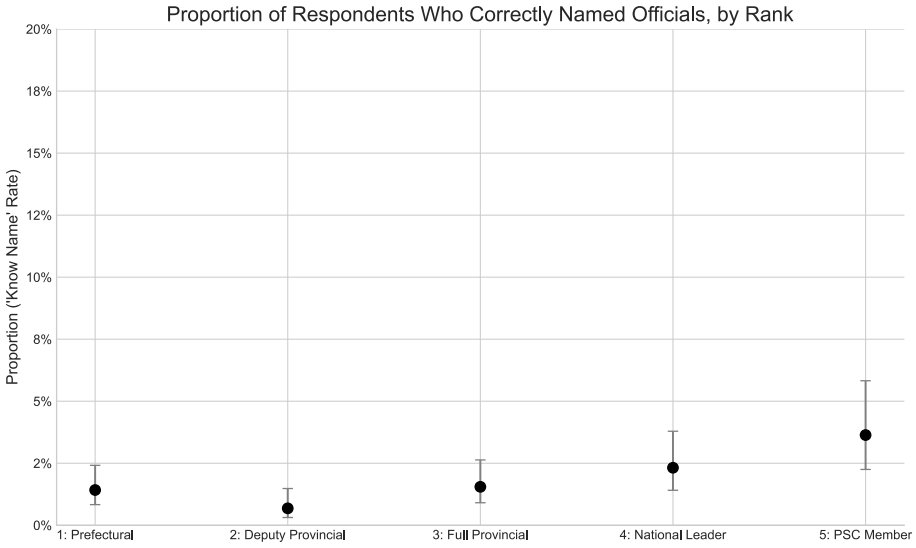



Figure B.5: Proportion of Respondents Correctly Identifying Officials, by Rank

Notes: The figure displays the proportion of respondents who correctly provided the name or position of the official. Dots represent the point estimate, and the vertical bars represent 95% confidence intervals.

Table B.2: Summary Statistics of Recognition Survey Respondents

| **Variable** | **Unweighted** | **Weighted** |
|---|---|---|
| *Panel A: Recognition Performance* | | |
| Correctly Identified an Official (%) | 2.2% | 2.5% |
| *Panel B: Demographics* | | |
| Mean Education Level | 3.60 | 3.72 |
| Male (%) | 65.6% | 60.6% |
| *Panel C: Media Habits* | | |
| Primary Media: Social Media (%) | 55.9% | 62.3% |
| Primary Media: Television (%) | 43.5% | 37.6% |
| *Panel D: News Preference* | | |
| Preferred News: Political (%) | 21.0% | 22.1% |
| Preferred News: Social/Lifestyle (%) | 9.7% | 6.0% |
| Total Number of Respondents | 186 | |

Notes: This table displays the summary statistics for the 186 unique respondents in the recognition survey. The 'Unweighted' column shows the simple averages from the raw survey data. The 'Weighted' column shows statistics after applying post-stratification weights to match the provincial distribution of the original human raters from the main study. Mean Education Level is measured on a 1-6 scale, where 1 = Middle School or less, 2 = High School, 3 = Junior College, 4 = Bachelor, 5 = Master, and 6 = PhD.

## B.3 Model Performance

Table B.3: Predictive Accuracy of Facial Traits Across Models

| Model | MSE | MAE | RMSE |
|---|---|---|---|
| **Perceived Attractiveness** | | | |
| AlexNet | 0.69 | 0.79 | 0.83 |
| DenseNet121 | 0.46 | 0.57 | 0.68 |
| DenseNet201 | 0.65 | 0.72 | 0.81 |
| ResNet18 | 0.68 | 0.79 | 0.82 |
| ResNet152 | 0.55 | 0.67 | 0.74 |
| **ResNeXt50** | 0.47 | 0.60 | 0.69 |
| ResNeXt101_32x8d | 0.67 | 0.75 | 0.82 |
| ViT | 0.78 | 0.84 | 0.88 |
| **Perceived Competence** | | | |
| AlexNet | 0.25 | 0.41 | 0.50 |
| DenseNet121 | 0.21 | 0.38 | 0.46 |
| DenseNet201 | 0.20 | 0.36 | 0.45 |
| ResNet18 | 0.24 | 0.41 | 0.49 |
| ResNet152 | 0.25 | 0.41 | 0.50 |
| **ResNeXt50** | 0.25 | 0.42 | 0.50 |
| ResNeXt101_32x8d | 0.23 | 0.39 | 0.48 |
| ViT | 0.29 | 0.44 | 0.53 |
| **Perceived Trustworthiness** | | | |
| AlexNet | 0.20 | 0.37 | 0.45 |
| DenseNet121 | 0.24 | 0.40 | 0.49 |
| DenseNet201 | 0.28 | 0.41 | 0.53 |
| ResNet18 | 0.20 | 0.36 | 0.45 |
| ResNet152 | 0.21 | 0.38 | 0.46 |
| **ResNeXt50** | 0.22 | 0.37 | 0.47 |
| ResNeXt101_32x8d | 0.20 | 0.36 | 0.45 |
| ViT | 0.24 | 0.40 | 0.49 |
| **Perceived Aggressiveness** | | | |
| AlexNet | 0.17 | 0.33 | 0.41 |
| DenseNet121 | 0.21 | 0.36 | 0.46 |
| DenseNet201 | 0.22 | 0.38 | 0.47 |
| ResNet18 | 0.17 | 0.32 | 0.41 |
| ResNet152 | 0.24 | 0.38 | 0.49 |
| **ResNeXt50** | 0.20 | 0.36 | 0.45 |
| ResNeXt101_32x8d | 0.19 | 0.35 | 0.44 |
| ViT | 0.17 | 0.33 | 0.41 |

***Note***: This table provides predictive accuracy metrics (MSE, MAE, RMSE) for different models (AlexNet, DenseNet121, ResNet18, ResNeXt50, DenseNet201, ResNet152, ResNeXt101_32x8d, ViT) across four facial traits: Perceived attractiveness, Perceived competence, Perceived trustworthiness, and Perceived aggressiveness.

## B.4 Supplementary Validation on a Curated Dataset

We conducted an additional validaiton test by sampling another 160 images. This set was constructed through a stratified random sampling method that slightly oversamples photos rated at the top and bottom 10%. This is to ensure that there is sufficient variation in human rating in a relative small sample. We then recruited a new group of 122 human raters to provide scores for the four traits on these 160 images.

Figure B.6 presents the scatter plots of the new human ratings against the model's original predictions for this curated dataset. The results show a marked improvement in the correlation between human and machine ratings. The Pearson's r coefficients are substantially higher than those from the cross-validation, ranging from $r = 0.397$ (Trustworthiness) to $r = 0.502$ (Aggressiveness), and all are highly statistically significant ($p < 0.001$).

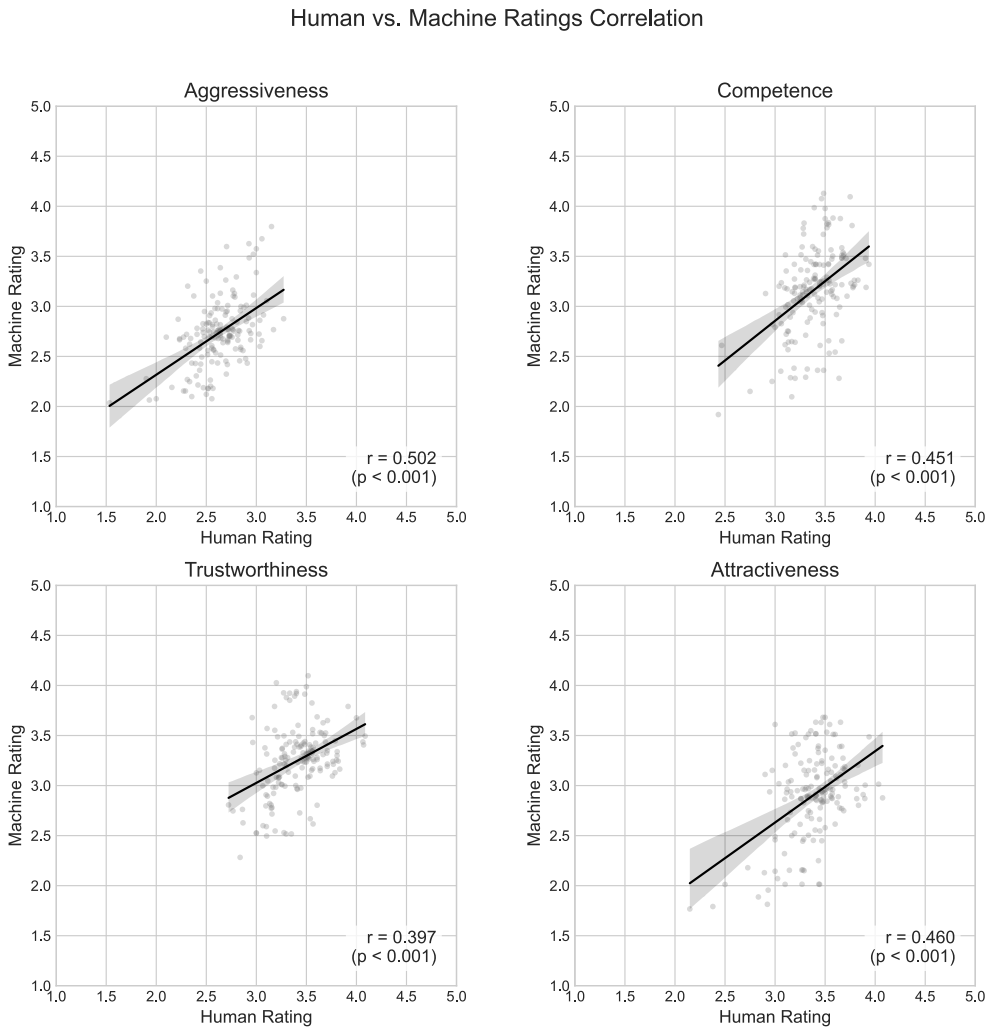
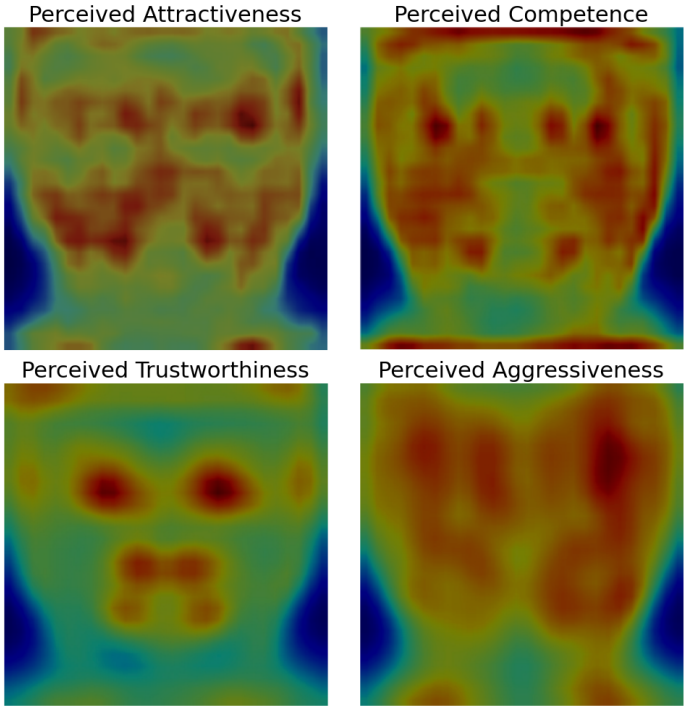



Figure B.6: Human vs. Machine Ratings on a Curated 160-Image Set

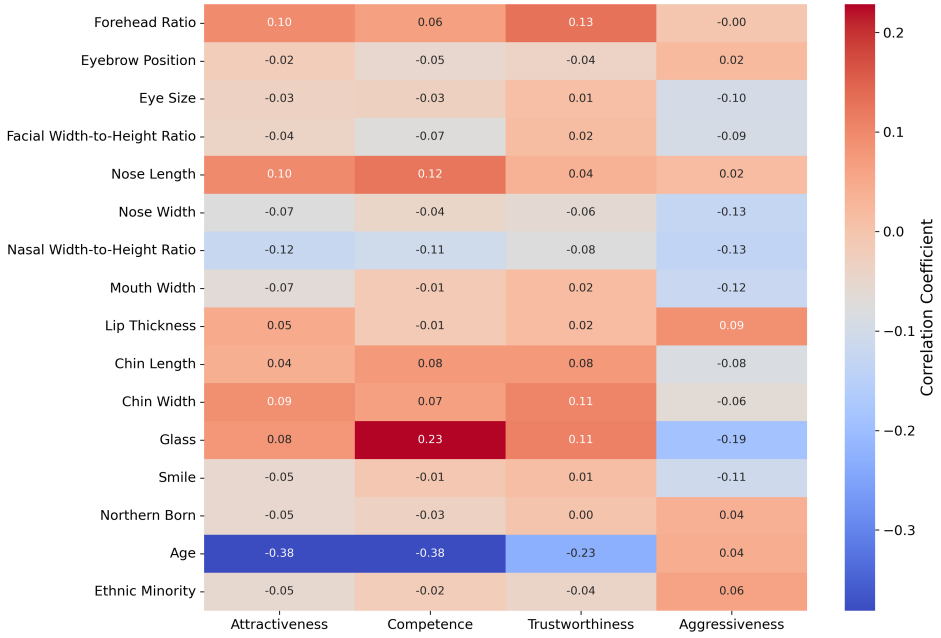
## B.5 Output Validation

Figure B.7: Contributions of Facial Regions to Trait Perceptions

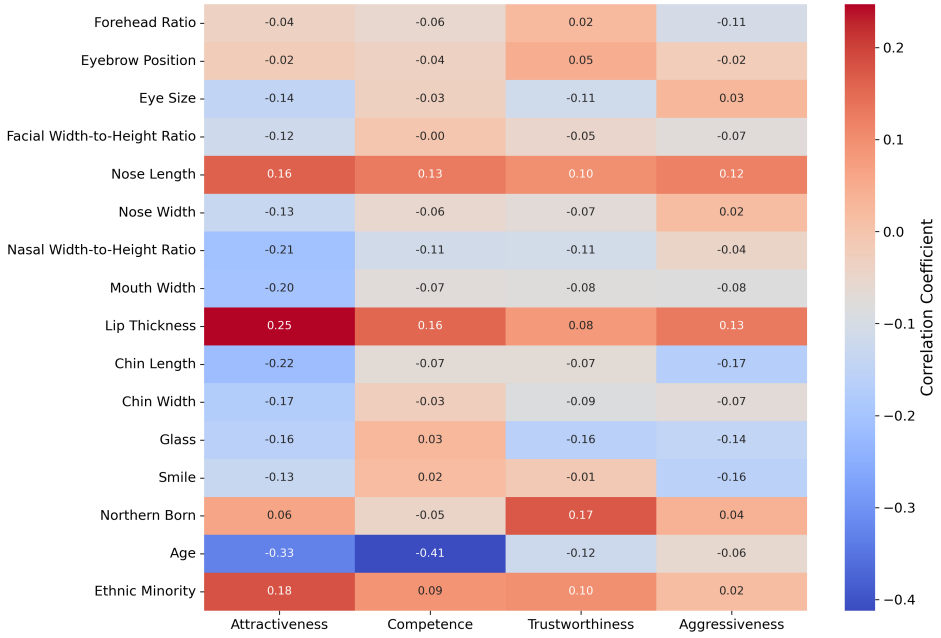



***Note:*** This figure illustrates the contributions of different facial regions to the machine-based ratings of the four facial traits. The facial heat maps are produced by Gradient-weighted Class Activation Mapping (Grad-CAM), a visualization technique for interpreting deep-learning models (Selvaraju et al. 2017). Regions highlighted in red indicate significant contributions to the prediction of the trait, while blue regions indicate minimal contributions. We see that inferences of attractiveness are primarily drawn from features in the central facial area, including the eyes, nose, and cheeks; perceptions of competence are most heavily influenced by the eyes, eyebrows, and jawline; evaluations of trustworthiness focus on the eye and mouth; finally, perceptions of aggressiveness rely on diffuse cues from the entire face, including the forehead, eyebrows, and broader lower facial regions. These patterns are broadly consistent with existing research on how humans infer traits from facial cues (e.g., Oosterhof and Todorov 2008; Vernon et al. 2014), suggesting that the machine-based algorithm replicates the processes underlying human perceptions of facial traits with reasonable accuracy and realism.

Figure B.8: Relationship between Objective Facial Features and Subjective Perceptions




Male




Female

Figure B.9: Examples of Photos Receiving Extreme Values in Automated Rating

Perceived attractiveness: highest

Perceived attractiveness: lowest

(a) Attractiveness

Perceived intelligence: highest

Perceived intelligence: lowest

(b) Competence

Perceived trustworthiness: highest

Perceived trustworthiness: lowest

(c) Trustworthiness

Perceived aggressiveness: highest

Perceived aggressiveness: lowest

(d) Aggressiveness

***Note:*** This figure shows the actual photos that receive the highest and lowest ratings for each trait. The rating is assigned by the ResNext-50 model with no human intervention.

Table B.4: Correlation Matrix of Perceived Facial Traits (Human Rating)

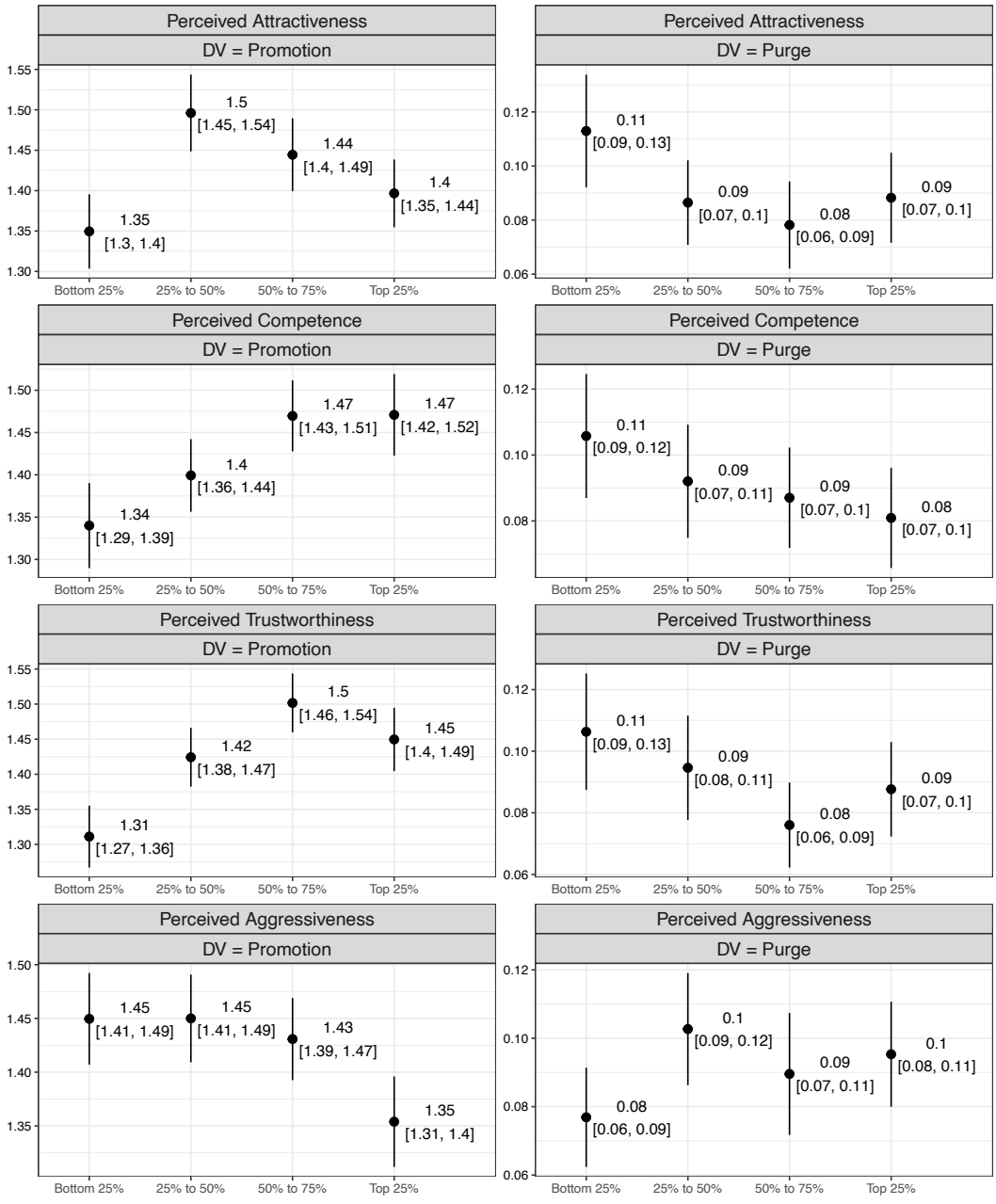
| | Perceived Attractiveness | Perceived Competence | Perceived Trustworthiness | Perceived Aggressiveness |
|---|---|---|---|---|
| Perceived Attractiveness | 1.00 | 0.42 | 0.33 | 0.08 |
| Perceived Competence | 0.41 | 1.00 | 0.32 | 0.02 |
| Perceived Trustworthiness | 0.33 | 0.32 | 1.00 | -0.002 |
| Perceived Aggressiveness | 0.08 | 0.02 | -0.002 | 1.00 |

Table B.5: Correlation Matrix of Perceived Facial Traits (Automated Rating)

| | Perceived Attractiveness | Perceived Competence | Perceived Trustworthiness | Perceived Aggressiveness |
|---|---|---|---|---|
| Perceived Attractiveness | 1.00 | 0.43 | 0.33 | -0.03 |
| Perceived Competence | 0.43 | 1.00 | 0.38 | -0.10 |
| Perceived Trustworthiness | 0.33 | 0.38 | 1.00 | -0.11 |
| Perceived Aggressiveness | -0.03 | -0.10 | -0.11 | 1.00 |

# C Additional Information on Robustness Checks

Figure B.1: Estimated Rank and Purge Probabilities Using Rating Quartiles




***Note:*** This figures shows the predicted rank and probability of purge for officials at different quartiles of facial ratings. The left and right columns are based on results from two separate regressions with different dependent variables (maximum rank and purge). In both regressions, we use the quartiles of facial ratings as the key independent variables. The rest of the specifications are identical to Models 5 and 10 of Table 1, respectively.

Table C.1: Robustness: Using Median Ratings Instead of Mean Ratings

| | DV: Political Rank | | | | | DV: Purge | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Perceived attractiveness (median) | 0.202** | | | | 0.084 | -0.059** | | | | -0.024 |
| | (0.053) | | | | (0.056) | (0.022) | | | | (0.026) |
| Perceived trustworthiness (median) | | 0.255** | | | 0.146* | | -0.078** | | | -0.051* |
| | | (0.059) | | | (0.065) | | (0.024) | | | (0.026) |
| Perceived competence (median) | | | 0.315** | | 0.227** | | | -0.080** | | -0.057* |
| | | | (0.060) | | (0.064) | | | (0.022) | | (0.025) |
| Perceived aggressiveness (median) | | | | -0.195** | -0.165** | | | | -0.005 | -0.013 |
| | | | | (0.064) | (0.063) | | | | (0.024) | (0.025) |
| Female | 0.210** | 0.221** | 0.227** | 0.215** | 0.214** | -0.033* | -0.036* | -0.038* | -0.037* | -0.036* |
| | (0.051) | (0.051) | (0.051) | (0.052) | (0.051) | (0.015) | (0.016) | (0.016) | (0.016) | (0.015) |
| Han ethnicity | -0.043 | -0.041 | -0.044 | -0.043 | -0.042 | -0.008 | -0.009 | -0.008 | -0.008 | -0.009 |
| | (0.055) | (0.054) | (0.054) | (0.054) | (0.054) | (0.020) | (0.020) | (0.021) | (0.021) | (0.020) |
| College degree | 0.453* | 0.486* | 0.462* | 0.478* | 0.480* | 0.136** | 0.126** | 0.133** | 0.132** | 0.131** |
| | (0.191) | (0.189) | (0.186) | (0.187) | (0.187) | (0.027) | (0.028) | (0.029) | (0.028) | (0.028) |
| Graduate degree | 0.599** | 0.637** | 0.609** | 0.629** | 0.626** | 0.153** | 0.142** | 0.150** | 0.149** | 0.149** |
| | (0.197) | (0.196) | (0.192) | (0.194) | (0.193) | (0.027) | (0.028) | (0.028) | (0.028) | (0.028) |
| Face quality (median) | 0.005** | 0.005** | 0.005** | 0.005** | 0.004** | -0.001* | -0.001+ | -0.001+ | -0.001* | -0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Blurriness (median) | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.27 | 0.27 | 0.27 | 0.26 | 0.27 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 |
| Observations | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 |

***Note***: This table presents the estimated effects of facial features on two sets of career outcomes (as of 2022): maximum political rank achieved and purge. For each official, we use the median rating of all his/her collected photos instead of the mean. The specification is otherwise identical to Table C.3. FE = fixed effects. Standard errors clustered at the birth city level are reported in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

Table C.2: Baseline Results with Facial Similarity Controls

| | DV: Political Rank | | | | | DV: Purge | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Perceived attractiveness | 0.268** | | | | 0.061 | -0.074** | | | | -0.027 |
| | (0.055) | | | | (0.060) | (0.024) | | | | (0.028) |
| Perceived competence | | 0.431** | | | 0.287** | | -0.096** | | | -0.063* |
| | | (0.064) | | | (0.069) | | (0.021) | | | (0.025) |
| Perceived trustworthiness | | | 0.414** | | 0.262** | | | -0.096** | | -0.060* |
| | | | (0.065) | | (0.071) | | | (0.026) | | (0.028) |
| Perceived aggressiveness | | | | -0.287** | -0.230** | | | | 0.011 | -0.002 |
| | | | | (0.069) | (0.068) | | | | (0.026) | (0.027) |
| Similarity to Xi Jinping | -0.002 | -0.002 | -0.003 | -0.002 | -0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | (0.004) | (0.003) | (0.004) | (0.004) | (0.004) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Similarity to Hu Jintao | 0.008** | 0.008** | 0.008** | 0.008* | 0.008** | 0.000 | -0.000 | 0.000 | 0.000 | -0.000 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Similarity to Jiang Zemin | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | -0.002 | -0.002 | -0.002 | -0.002 | -0.002 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Female | 0.287** | 0.299** | 0.292** | 0.280** | 0.281** | -0.042* | -0.046* | -0.044* | -0.044* | -0.044* |
| | (0.058) | (0.057) | (0.057) | (0.058) | (0.057) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) |
| Han ethnicity | -0.059 | -0.052 | -0.057 | -0.058 | -0.054 | -0.005 | -0.007 | -0.006 | -0.006 | -0.007 |
| | (0.057) | (0.056) | (0.055) | (0.055) | (0.055) | (0.021) | (0.021) | (0.021) | (0.021) | (0.021) |
| College degree | 0.488* | 0.487* | 0.522** | 0.517** | 0.516** | 0.126** | 0.126** | 0.118** | 0.122** | 0.123** |
| | (0.194) | (0.190) | (0.190) | (0.190) | (0.189) | (0.028) | (0.029) | (0.027) | (0.028) | (0.028) |
| Graduate degree | 0.635** | 0.634** | 0.674** | 0.670** | 0.662** | 0.143** | 0.143** | 0.134** | 0.138** | 0.140** |
| | (0.200) | (0.196) | (0.198) | (0.196) | (0.196) | (0.027) | (0.029) | (0.027) | (0.028) | (0.028) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.25 | 0.26 | 0.26 | 0.25 | 0.27 | 0.13 | 0.13 | 0.13 | 0.12 | 0.13 |
| Observations | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 |

***Note***: This table presents the estimated effects of facial features on two sets of career outcomes (as of 2022): maximum political rank achieved and purge. *Similarity to Xi Jinping*, *Similarity to Hu Jintao*, and *Similarity to Jiang Zemin* are variables that are constructed to measure the facial similarity between an official and the three consecutive top leaders following the approach of Schroff, Kalenichenko, and Philbin (2015). FE = fixed effects. Standard errors clustered at the birth city level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$

Table C.3: Baseline Results with Image Quality Controls: Promotion and Purge

| | DV: Political Rank | | | | | DV: Purge | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Perceived attractiveness | 0.230** | | | | 0.058 | -0.062** | | | | -0.026 |
| | (0.057) | | | | (0.061) | (0.024) | | | | (0.028) |
| Perceived competence | | | 0.375** | | 0.259** | | | -0.082** | | -0.057* |
| | | | (0.065) | | (0.069) | | | (0.022) | | (0.025) |
| Perceived trustworthiness | | 0.346** | | | 0.213** | | -0.073** | | | -0.041 |
| | | (0.065) | | | (0.072) | | (0.026) | | | (0.029) |
| Perceived aggressiveness | | | | -0.260** | -0.217** | | | | 0.001 | -0.008 |
| | | | | (0.069) | (0.068) | | | | (0.026) | (0.027) |
| Female | 0.212** | 0.221** | 0.226** | 0.213** | 0.213** | -0.032* | -0.035* | -0.036* | -0.035* | -0.035* |
| | (0.051) | (0.051) | (0.051) | (0.052) | (0.051) | (0.015) | (0.015) | (0.015) | (0.015) | (0.015) |
| Han ethnicity | -0.043 | -0.043 | -0.038 | -0.043 | -0.041 | -0.009 | -0.010 | -0.010 | -0.010 | -0.010 |
| | (0.056) | (0.055) | (0.055) | (0.054) | (0.054) | (0.020) | (0.020) | (0.021) | (0.021) | (0.020) |
| College degree | 0.477* | 0.509** | 0.479* | 0.502** | 0.505** | 0.135** | 0.127** | 0.134** | 0.132** | 0.132** |
| | (0.190) | (0.189) | (0.187) | (0.187) | (0.187) | (0.028) | (0.028) | (0.029) | (0.029) | (0.029) |
| Graduate degree | 0.624** | 0.659** | 0.625** | 0.653** | 0.651** | 0.153** | 0.144** | 0.152** | 0.149** | 0.150** |
| | (0.196) | (0.196) | (0.193) | (0.193) | (0.193) | (0.028) | (0.028) | (0.029) | (0.029) | (0.028) |
| Face quality | 0.004** | 0.003** | 0.004** | 0.004** | 0.003** | -0.001** | -0.001** | -0.001** | -0.002** | -0.001** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Blurriness | -0.001 | -0.002 | -0.001 | -0.001 | -0.002 | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.26 | 0.26 | 0.26 | 0.26 | 0.27 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 |
| Observations | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 | 4071 |

***Note***: This table presents the estimated effects of facial features on two sets of career outcomes (as of 2022): maximum political rank achieved and purge. *Face quality* and *Blurriness* are two measures of image quality generated by the Face++ API. FE = fixed effects. Standard errors clustered at the birth city level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$

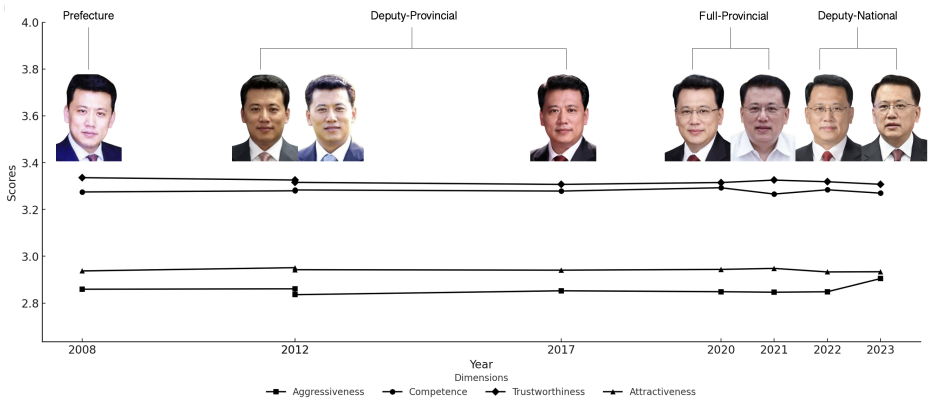
Table C.4: Official Portraits vs. Meeting and Event Photos

| | Promotion | | Purge | |
|---|---|---|---|---|
| | (1) Head shot | (2) Meeting | (3) Head shot | (4) Meeting |
| Perceived attractiveness | 0.083$^{+}$ | 0.066 | -0.020 | -0.029 |
| | (0.043) | (0.058) | (0.019) | (0.024) |
| Perceived competence | 0.193$^{**}$ | 0.173$^{**}$ | -0.022 | -0.041$^{+}$ |
| | (0.045) | (0.062) | (0.019) | (0.024) |
| Perceived trustworthiness | 0.138$^{**}$ | 0.163$^{**}$ | -0.004 | -0.034 |
| | (0.051) | (0.058) | (0.020) | (0.025) |
| Perceived aggressiveness | -0.158$^{**}$ | -0.128$^{+}$ | -0.008 | 0.025 |
| | (0.046) | (0.069) | (0.018) | (0.025) |
| Female | 0.229$^{**}$ | 0.199$^{**}$ | -0.038$^{*}$ | -0.036$^{*}$ |
| | (0.054) | (0.053) | (0.016) | (0.016) |
| Han ethnicity | -0.044 | -0.043 | -0.009 | -0.011 |
| | (0.056) | (0.054) | (0.022) | (0.021) |
| College degree | 0.435 | 0.545$^{**}$ | 0.143$^{**}$ | 0.120$^{**}$ |
| | (0.339) | (0.205) | (0.039) | (0.031) |
| Graduate degree | 0.580$^{+}$ | 0.686$^{**}$ | 0.163$^{**}$ | 0.138$^{**}$ |
| | (0.347) | (0.212) | (0.038) | (0.031) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.29 | 0.27 | 0.14 | 0.13 |
| Observations | 3771 | 3920 | 3771 | 3920 |

***Note***: This table presents the estimated effects of facial features on promotions and purges. We use facial feature ratings generated solely from official portraits (Models 1 and 3) and meeting photos (Models 2 and 4). Demographic covariates include *Gender* and *Han ethnicity*, and *Education level*. FE = fixed effects.

Standard errors clustered at the birth city level are reported in parentheses. $^{+}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

Figure B.2: Temporal Stability in Facial Rating for Selected Figures




Yuan Jiajun

Table C.5: Effects of Promotion on Facial Features

| | DV: Facial Traits Rating | | | |
|---|---|---|---|---|
| | (1) Perceived attractiveness | (2) Perceived trustworthiness | (3) Perceived competence | (4) Perceived aggressiveness |
| Photos taken after Politburo | 0.076 | -0.019 | 0.024 | -0.008 |
| | (0.067) | (0.051) | (0.066) | (0.032) |
| Age | -0.007 | 0.002 | -0.014** | 0.000 |
| | (0.005) | (0.004) | (0.005) | (0.003) |
| Blurriness | 0.008* | 0.007** | 0.008** | 0.001 |
| | (0.003) | (0.002) | (0.002) | (0.003) |
| Face quality | 0.004* | 0.003* | 0.004** | -0.001 |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Individual FE | ✓ | ✓ | ✓ | ✓ |
| Observations | 352 | 352 | 352 | 352 |
| Adj. R-squared | 0.204 | 0.209 | 0.276 | 0.261 |

***Note***: This table presents the estimated effects of how promotion and age affect facial features of Politburo members. The sample includes 352 photos for 37 non-military members in the 19th and 20th Central Committee. *Photos taken after Politburo* is a binary indicator that takes the value of 1 if the photo was taken after the person had been promoted to the Politburo, and 0 otherwise. *Age* is a person's actual age at that time the photo was taken. We also include two Image quality controls derived from the Face++ API: *Face quality*, which assesses general image quality, and *Blurness*, which evaluates the level of blur in the image. FE = fixed effects. Standard errors are clustered at the individual level and reported in parentheses. $^{+}$ $p < 0.1$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$.

Table C.6: Effects of Facial Features on City Leaders' Economic and Fiscal Performance

| | DV: GDP Growth at $t+2$ | | | | | DV: Fiscal Revenue Growth at $t+2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Perceived attractiveness | -0.022 | | | | -0.015 | -0.011 | | | | 0.004 |
| | (0.023) | | | | (0.026) | (0.050) | | | | (0.053) |
| Perceived competence | | -0.037 | | | -0.038 | | -0.056 | | | -0.065 |
| | | (0.027) | | | (0.031) | | (0.051) | | | (0.055) |
| Perceived trustworthiness | | | 0.004 | | 0.024 | | | 0.008 | | 0.033 |
| | | | (0.024) | | (0.027) | | | (0.047) | | (0.048) |
| Perceived aggressiveness | | | | 0.006 | 0.002 | | | | 0.056 | 0.048 |
| | | | | (0.030) | (0.031) | | | | (0.059) | (0.061) |
| Female | 0.044 | 0.042 | 0.042 | 0.042 | 0.043 | 0.054 | 0.054 | 0.052 | 0.055 | 0.055 |
| | (0.023) | (0.022) | (0.023) | (0.023) | (0.023) | (0.033) | (0.033) | (0.033) | (0.033) | (0.034) |
| Han ethnicity | 0.027 | 0.028 | 0.027 | 0.027 | 0.027 | 0.038 | 0.037 | 0.037 | 0.037 | 0.035 |
| | (0.030) | (0.030) | (0.030) | (0.030) | (0.030) | (0.044) | (0.045) | (0.044) | (0.044) | (0.045) |
| College degree | -0.022 | -0.022 | -0.022 | -0.023 | -0.022 | -0.020 | -0.020 | -0.020 | -0.021 | -0.022 |
| | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.043) | (0.044) | (0.043) | (0.043) | (0.044) |
| Graduate degree | -0.029 | -0.028 | -0.029 | -0.029 | -0.028 | -0.030 | -0.030 | -0.031 | -0.031 | -0.031 |
| | (0.017) | (0.017) | (0.017) | (0.017) | (0.017) | (0.039) | (0.040) | (0.039) | (0.039) | (0.040) |
| Birth year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Birth city FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.53 | 0.53 | 0.53 | 0.53 | 0.53 |
| Observations | 2016 | 2016 | 2016 | 2016 | 2016 | 1815 | 1815 | 1815 | 1815 | 1815 |

***Note***: This table presents the estimated effects of facial features on city leaders' economic and fiscal performance at $t+2$. The $t+2$ GDP and fiscal revenue growth rates are calculated as the percentage change relative to the first year of the city leader's tenure. FE = fixed effects. Standard errors clustered at the birth city level are reported in parentheses. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

Table C.7: Pooled Model with Interaction Terms (By Step)

| | DV: Promoted (1 = yes) |
|---|---|
| Promotion step = 2 | 0.523 |
| | (0.361) |
| Promotion step = 3 | -1.119** |
| | (0.362) |
| Perceived attractiveness | 0.033 |
| | (0.032) |
| Perceived competence | 0.108** |
| | (0.034) |
| Perceived trustworthiness | 0.010 |
| | (0.034) |
| Perceived aggressiveness | -0.051 |
| | (0.038) |
| Attractiveness × Step 2 | -0.101 |
| | (0.071) |
| Attractiveness × Step 3 | -0.060 |
| | (0.079) |
| Competence × Step 2 | -0.038 |
| | (0.079) |
| Competence × Step 3 | 0.162* |
| | (0.096) |
| Trustworthiness × Step 2 | 0.177** |
| | (0.072) |
| Trustworthiness × Step 3 | 0.146 |
| | (0.092) |
| Aggressiveness × Step 2 | -0.243** |
| | (0.081) |
| Aggressiveness × Step 3 | 0.063 |
| | (0.087) |
| Female | 0.073*** |
| | (0.020) |
| Han ethnicity | -0.028 |
| | (0.021) |
| College degree | 0.158*** |
| | (0.027) |
| Graduate degree | 0.244*** |
| | (0.026) |
| Birth year FE | ✓ |
| Birth city FE | ✓ |
| Observations | 5,996 |
| R-squared | 0.124 |

**Note:** This table reports the results of a pooled regression model where each observation corresponds to an official at a given promotion step. The model includes interactions between each facial trait and the indicator for the promotion step. Standard errors clustered at the birth city level are shown in parentheses. FE = fixed effects. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table C.8: Step-Specific Marginal Effects from the Pooled Interaction Model

| | **Effect on Promotion Probability** | | |
|---|---|---|---|
| | **Step 1** (Pref. to Deputy) | **Step 2** (Deputy to Full) | **Step 3** (Full to National) |
| **Facial Traits** | | | |
| Attractiveness | 0.033 | -0.068 | -0.027 |
| | (0.032) | (0.071) | (0.079) |
| Competence | 0.108*** | 0.070 | 0.270*** |
| | (0.034) | (0.079) | (0.096) |
| Trustworthiness | 0.010 | 0.187*** | 0.156* |
| | (0.034) | (0.072) | (0.092) |
| Aggressiveness | -0.051 | -0.294*** | 0.012 |
| | (0.038) | (0.081) | (0.087) |
| **Control Variables** | | | |
| Female | | 0.073*** | |
| | | (0.020) | |
| Han Ethnicity | | -0.028 | |
| | | (0.021) | |
| College Degree | | 0.158*** | |
| | | (0.027) | |
| Graduate Degree | | 0.244*** | |
| | | (0.026) | |
| Observations | | 5,996 | |
| Birth Year FE | | ✓ | |
| Birth City FE | | ✓ | |

**Note:** This table presents the estimated total effect of each facial trait on promotion at three career steps. Effects for facial traits are calculated from a single pooled regression model with interactions. Effects for control variables are from the same model and are constant across steps. Standard errors clustered at the birth city level are shown in parentheses. FE = fixed effects. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table C.9: Pooled Model with Interaction Terms (By Gender)

| | **Promotion** (1) | **Purge** (2) |
|---|---|---|
| Female (=1) | -2.044 | 0.832 |
| | (1.714) | (0.543) |
| Perceived attractiveness | 0.072 | -0.020 |
| | (0.064) | (0.029) |
| Perceived competence | 0.289** | -0.063* |
| | (0.070) | (0.025) |
| Perceived trustworthiness | 0.237** | -0.054 |
| | (0.073) | (0.030) |
| Perceived aggressiveness | -0.265** | -0.003 |
| | (0.071) | (0.028) |
| Female × Attractiveness | -0.242 | -0.101 |
| | (0.313) | (0.087) |
| Female × Competence | -0.018 | -0.080 |
| | (0.354) | (0.093) |
| Female × Trustworthiness | 0.454 | -0.110 |
| | (0.329) | (0.087) |
| Female × Aggressiveness | 0.575 | 0.003 |
| | (0.464) | (0.117) |
| Han ethnicity | -0.056 | -0.010 |
| | (0.054) | (0.020) |
| College degree | 0.514** | 0.126** |
| | (0.189) | (0.028) |
| Graduate degree | 0.662** | 0.144** |
| | (0.196) | (0.028) |
| Constant | -0.248 | 0.386** |
| | (0.386) | (0.127) |
| Birth year FE | ✓ | ✓ |
| Birth city FE | ✓ | ✓ |
| Observations | 4010 | 4010 |

**Note:** This table reports the results of a pooled regression model that includes interaction terms between each facial trait and an indicator for female candidates. Unlike the subsample analysis in the main text, this model estimates gender heterogeneity by allowing the effects of facial traits to vary systematically between male and female officials within a single unified specification. Standard errors clustered at the birth-city level are reported in parentheses. FE = fixed effects. $^{*}p < 0.05$, $^{**}p < 0.01$.

Table C.10: Gender-Specific Marginal Effects from the Pooled Interaction Model

| | DV: Promotion (maxrank) | | DV: Purge | |
|---|---|---|---|---|
| | **Male** | **Female** | **Male** | **Female** |
| **Facial Traits** | | | | |
| Attractiveness | 0.072 | -0.171 | -0.020 | -0.121 |
| | (0.064) | (0.313) | (0.029) | (0.087) |
| Competence | 0.289*** | 0.271 | -0.063** | -0.143 |
| | (0.070) | (0.354) | (0.025) | (0.093) |
| Trustworthiness | 0.237*** | 0.700** | -0.054* | -0.164* |
| | (0.073) | (0.329) | (0.030) | (0.087) |
| Aggressiveness | -0.265*** | 0.310 | -0.003 | 0.000 |
| | (0.071) | (0.464) | (0.028) | (0.117) |
| **Control Variables** | | | | |
| Han Ethnicity | -0.056 | | -0.010 | |
| | (0.054) | | (0.020) | |
| College Degree | 0.514*** | | 0.126*** | |
| | (0.189) | | (0.028) | |
| Graduate Degree | 0.662*** | | 0.144*** | |
| | (0.196) | | (0.028) | |
| Observations | 4,010 | | 4,010 | |
| R-squared | 0.251 | | 0.121 | |
| Birth Year FE | Yes | | Yes | |
| Birth City FE | Yes | | Yes | |

**Note:** This table reports gender-specific marginal effects implied by two pooled regression models with trait-by-gender interactions (one for promotion and one for purge). Male effects correspond to the main coefficients; female effects equal the sum of the main coefficient and the interaction term. Control variables are constant across genders. Standard errors are clustered at the birth city level. FE = fixed effects. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

# D Additional Information on Conjoint Experiment

## D.1 Experiment Interface

We first present respondents with a brief prompt about the task and then show them eight successive pairs of candidate profiles. The wording of the prompt and the examples of the profiles are as follows. The two photos are matched on attractiveness, competence, and aggressiveness, but the one on the right has higher trustworthiness rating than the one on the left. The colored texts highlight the randomized components of the conjoint experiment (The subjects only saw black texts in the actual experiments).

**Prompt** "假设您是某政府机关单位的主要负责同志，即将对以下两位干部进行考察，并选择提拔其中一位，以下是这两名干部的一些基本信息和简历照片。综合他们的背景，以及您看到他们相貌的第一印象，您觉得哪位更有可能被提拔？" ("Suppose you are the main decision maker in a government agency and are about to evaluate the following two cadres, choosing one for promotion. Below is the photo and background information of these two candidates. Based on your reading of their profiles and impression of their appearance, which one do you think is more likely to be promoted?")

干部1：1977年出生，本科学历，曾在基层政府任职

Candidate 1 was born in 1977, has a bachelor's degree, and has worked in grassroots government

干部2：1973年出生，博士研究生，曾担任上级领导秘书

Candidate 2 was born in 1973, has a Ph.D degree, and has worked as a personal aide to a senior leader

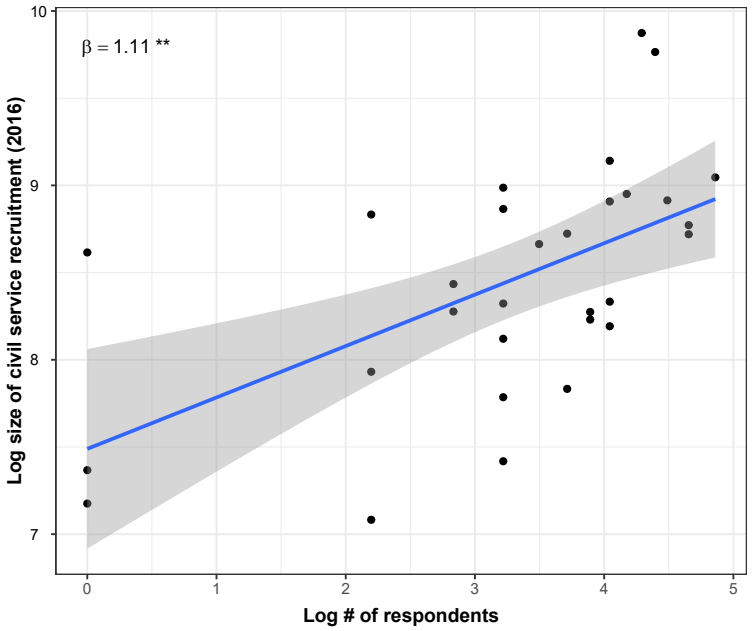
## D.2 Validation and Numerical Results

Table D.1: Numerical Results for Conjoint Experiment

| | DV: Profile Selected for Promotion | | | |
|---|---|---|---|---|
| | (1) Perceived attractiveness | (2) Perceived competence | (3) Perceived trustworthiness | (4) Perceived aggressiveness |
| Face with higher rating | -0.017 | 0.249** | 0.312** | -0.334** |
| | (0.055) | (0.053) | (0.051) | (0.056) |
| Birth year (baseline = 1973) | | | | |
| - 1975 | -0.005 | 0.076 | 0.235* | -0.122 |
| | (0.107) | (0.096) | (0.099) | (0.101) |
| - 1977 | 0.072 | 0.063 | 0.113 | 0.154 |
| | (0.111) | (0.087) | (0.092) | (0.104) |
| - 1979 | 0.018 | 0.100 | 0.213* | 0.047 |
| | (0.117) | (0.103) | (0.095) | (0.103) |
| Education (baseline = bachelor's degree) | | | | |
| - Master's degree | 0.124 | 0.323** | 0.348** | 0.156+ |
| | (0.094) | (0.077) | (0.078) | (0.085) |
| - Ph.D degree | 0.375** | 0.564** | 0.309** | 0.406** |
| | (0.102) | (0.080) | (0.084) | (0.092) |
| Work experience (baseline = SOE) | | | | |
| - Grassroots government | 0.137 | 0.189* | 0.034 | 0.096 |
| | (0.110) | (0.094) | (0.101) | (0.102) |
| - Central government | 0.429** | 0.531** | 0.437** | 0.524** |
| | (0.123) | (0.100) | (0.099) | (0.110) |
| - Personal aide to senior leader | 0.454** | 0.480** | 0.365** | 0.324** |
| | (0.117) | (0.101) | (0.091) | (0.108) |
| Respondent-pair FE | ✓ | ✓ | ✓ | ✓ |
| R-squared | 0.12 | 0.28 | 0.25 | 0.24 |
| # of subjects | 159 | 159 | 159 | 159 |
| Observations | 636 | 636 | 636 | 636 |

***Note***: This table presents the numerical estimates from the conjoint experiment on civil servants. The units of observation are individual official profiles (photo + short biography). The dependent variable is whether a given profile is selected by a respondent for promotion in a pair. Respondent–pair fixed effects are included in all models and standard errors clustered at respondent level are reported in parentheses. SOE = state-owned enterprises. * $p < 0.05$, ** $p < 0.01$

Figure B.1: Representativeness of Conjoint Experiment Respondents




***Note***: This table presents the correlation between the regional origin of respondents in our conjoint sample and the actual regional distribution of government size in China. Each circle represents a province. The x-axis is the logged number of conjoint experiment respondents from that province, and the y-axis is the logged number of new civil servants recruited in 2016. The coefficient printed at the top left is from the bivariate regression between the two variables.