

Uncertainty in Bureaucracy

*Toward a Sociological Understanding
of Frontline Decision Making*



Nadine Raaphorst

Uncertainty in Bureaucracy

*Toward a Sociological Understanding
of Frontline Decision Making*

Nadine Raaphorst

© 2017 Nadine Raaphorst

ISBN: 9789462957015

Cover & layout: Kjell Betlem, Frameloos

Printing: ProefschriftMaken

This dissertation was funded by the Dutch Organization for Scientific Research (NWO), Vidi project ‘The other side of the gap: Do public officials trust citizens?’ (grant no. 452-11-011).

Uncertainty in Bureaucracy

Toward a Sociological Understanding
of Frontline Decision Making

Onzekerheid in bureaucratie

Naar een sociologisch begrip van besluitvorming
door contactambtenaren

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 12 oktober 2017 om 15.30 uur

Nadine Jasmijn Raaphorst
geboren te Kwintsheul

Doctoral Committee

Doctoral supervisors:

Prof.dr. S. Van de Walle

Prof.dr. S.M. Groeneveld

Other members:

Prof.dr. A.J. Steijn

Prof.dr. P. Mascini

Dr. G.S. Harrits

Table of contents

Table of contents

Acknowledgements	9
Chapter 1	13
1.1. Bureaucracy and uncertainty	13
1.2. Uncertainty in street-level work	14
1.3. Reducing uncertainty: Social network, signals and stereotypes	16
1.4. Research aim and general research question	16
1.5. Research approach and setting	17
1.6. Theoretical and methodological relevance	20
1.7. Societal relevance	23
1.8. Structure and outline of dissertation	24
Chapter 2	27
Abstract	27
2.1. Introduction	28
2.2. Review of literature	29
2.3. Research site	30
2.4. Methods	31
2.5. Findings	33
2.6. Conclusion and discussion	41
Chapter 3	45
Abstract	45
3.1. Introduction	46
3.2. The social context of frontline decision making	47
3.3. Description of research contexts	48
3.4. Methodology	50
3.5. Findings	50
3.6. Discussion	59
3.7. Conclusion and recommendations for future research	60
Chapter 4	63
Abstract	63
4.1. Introduction	64
4.2. Review of literature	65
4.3. Research site: The Dutch tax authorities	68
4.4. Respondent selection	69
4.5. Methods	70
4.6. Findings	71
4.7. Conclusion and discussion	78
Chapter 5	83
Abstract	83
5.1. Introduction	84

5.2. Stereotyping at the frontline	84
5.3. Theoretical framework	86
5.4. Dutch tax administration	87
5.5. Hypotheses	88
5.6. The policy-capturing study	90
5.7. Findings	92
5.8. Conclusion & discussion	96
Chapter 6	101
6.1. A study on bureaucratic uncertainty	101
6.2. Conclusions	103
6.3. General conclusion	106
6.4. Societal relevance	109
6.5. Discussion	110
6.6. Limitations of the study	114
6.7. Recommendations for future research	115
Bibliography	121
Appendices	129
Dutch summary	139
About the author	145
List of tables and figures	
Table 2.1. Coding table with number of stories	33
Table 2.2. Description of the kinds of uncertainty at play in frontline tax officials' work	40
Table 3.1. Description of the cases	51
Table 3.2. Summary of the main empirical findings	58
Table 4.1. Conceptualization of main concepts	67
Table 4.2. Double standards mechanism and resulting strictness of treatment	68
Table 4.3. Sources of signals (leftmost column) and examples of kinds of signals looked at to assess trustworthiness and untrustworthiness	72
Table 4.4. Status characteristics as framework leading to different interpretations	75
Figure 5.1. Conceptual model with hypotheses	88
Table 5.1. Descriptive statistics dependent variables	91
Table 5.2. Multilevel analyses of direct and interaction effects	94
Figure 5.2. Graph of interaction effect Dodging around question	95
* Level of education on 'appears okay'	

Acknowledgements

Completing this dissertation would not have been possible were it not for the incredible people who supported and motivated me the last four years. I want to express my gratitude to a number of people for their invaluable support, guidance and encouragement.

First and foremost, I would like to thank my supervisors, Steven Van de Walle and Sandra Groeneveld, for giving me the opportunity to pursue this Ph.D. project. Your guidance and wisdom were greatly appreciated. Steven, thanks for encouraging me to develop my own research ideas and academic network. I have learned a lot from your advice on strategic career decisions. I'm looking forward collaborating with you in the future. Sandra, you've been a great inspiration and support for me. During our meetings, you're always knowledgeable and precise. I've enjoyed and learned so much from our conversations about concepts, theories and developments in public administration research. I'm glad to be your colleague at the Leiden University now, so we can continue having these conversations and collaborating on joint papers.

I'd also like to thank a number of people from the Dutch tax administration without whom my fieldwork would not have been possible. I'm indebted to Theo Poolen, Driek Kamps, Wouter Smit, Rex Arendsen and Albert van Steenberg, who enabled the research project's execution. I'm furthermore grateful to the members of the advisory board – Hanskje Nagel, Janneke de Jonge, Hanneke Hermsen, Albert van Steenberg – for their constructive criticism and advice on my research proposals. I'm furthermore grateful to all the managers and respondents who were willing to participate to my research.

There is a number of people I have met throughout my Ph.D. I'd like to thank. Gitte, Marie and Mathilde, thanks for making my research stay in Aarhus unforgettable. I really appreciate and enjoyed the meetings and dinners we had. I would also like to thank Kim Loyens for the fruitful collaboration and Peter Hupe for letting me present my work in his panels at various conferences. A special thanks to Alicia Slooff for helping me in transcribing the interviews.

Besides being a serious business, writing a Ph.D. also meant having fun with colleagues. Shelena and Ziya, your crazy and funny stories in T17-46 made my working days so much more fun. Shelena, I already miss your questionable music suggestions (yes, sometimes I did listen to them). I really enjoyed working with you in the 'Vidi project' and hope we'll keep in touch. I'd also like to thank Wey, Ewald, Rianne, Roy, Nadine, Stéphane, Warda, William, Ingmar, Sanne, Stephan, Anne and Pablo for the entertaining lunch conversations, coffee breaks, conference dinners, and after work drinks. Thanks to my new colleagues at FGGA – Tanachia, Eduard, Machiel, Marieke, Petra, Bert, Valérie, Joris, Bernard, Chantal and Mette, to name just a few – for making me feel very welcome.

Some colleagues deserve extra mention because they have become my friends. Ilona, thanks for your friendship. I really appreciate the good conversations we've had during coffee breaks and dinners, and I value our shared music interest. I hope we'll visit many more concerts in the

Acknowledgements

future. Also thanks to Evelien and Anna; your critical and feminist take on the world is enlightening and very much needed! Evelien, I really appreciate our deep conversations which never seem to fit our one-hour lunch breaks. Dion, thank you for your friendship and your endless music and novel suggestions. The last 1,5 years of my Ph.D. wouldn't have been the same without the 'Totem'; Alette, Lieselot, Noortje and Tebbine, thanks for all the fun and support I desperately needed in finishing my dissertation. I look forward to all the trips we're going to make, from skiing holidays in La Clusaz to visiting Tebbine's future goat farm in Geneva. A special thanks to Lieselot; I value the advices you've given me, and thank you for getting me hooked to bouldering. A special thanks to Noortje too; I enjoyed our 8:30 A.M. coffee moments, your cheerfulness and your exaggerations, and appreciate the methodological help you've given me.

My friends outside academia also reminded me there's more to life than completing a dissertation; Marlies, Lisanne, Denise, Claudia, Debby and Marlou, I'm so happy having you as friends. I furthermore enjoyed the dinners and get-togethers with Sjakkeliën, Harry, Sven, Eric, Wilma, Joke, Leo, Henny and Michèl. Thank you for your interest in my research and the fun we have had.

Of course, I could not have done this Ph.D. without the unconditional support of my family. Ingrid and René, you're the best sister and brother I could ever ask for. Ing, I cherish our good conversations and I'm happy to share things with you that only you understand. René, my big 'little brother', I (generally) love your unannounced visits and jokes. Thanks for making me laugh when I needed it. Mom and dad, I dedicate this dissertation to you. You have always been there for me and supported me in my life decisions. You never held me back, and always encouraged me to follow my dreams. You taught me to work hard and to be kind, for which I'm deeply grateful.

Last but not least, Kjell, my dearest, words fail me to express my gratitude for having you in my life. I greatly value your continuous support, encouragement and love. Your ways of approaching things in life are uplifting and so meaningful to me. I'm looking forward to the future we have ahead of us. Thank you for always believing in me.

Rotterdam, July 2017

Chapter 1

Introduction

1.1. Bureaucracy and uncertainty

Bureaucracy, in its ideal typical form, has been depicted by Weber as ‘fundamentally the exercise of control on the basis of knowledge’ (1978, p. 339). As part of the process of bureaucratic rationalization, a means-end calculus became the organizational form of modern life. Through the ‘rational’ character of bureaucracy, Weber held, a high degree of calculability of results could be achieved. The pure form is characterized by ‘Precision, speed, unambiguity, knowledge of the files, continuity, discretion, unity, strict subordination, reduction of friction and of material and personal costs’ (Weber, 1978, p. 973). As such, the bureaucratic organization reduces uncertainty ‘both in its internal procedures and in its outputs’ (Gajduschek, 2003, p. 718; see also Groeneveld, 2016).

This reduction of uncertainty, it is held, is a precondition for the functioning of modern democratic systems, since it guarantees that the general interest is served, and, hence, is to the advantage of both rulers and citizen-clients (Gajduschek, 2003). Within a representative government, adherence to rules and hierarchy is seen as necessary to guarantee that goals and rules laid down by elected officials are actually carried out. A bureaucrat thereby is ‘subject to an impersonal order to which his actions are oriented’ (Weber, 1978, p. 330). As such, bureaucracy is seen as a powerful instrument for those at the top of an organization to organize authority relations, since outputs are calculable and processes controllable. Bureaucratic organizations are also considered to reduce uncertainty for citizen-clients, since the ‘spirit of formalistic impersonality’ (Weber, 1978, p. 340), ensures that everyone in the same situation is treated equally. Within this ideal typical view of bureaucracy, emotions such as hatred, affection and enthusiasm are viewed as threats to bureaucratic functioning, opening the door to arbitrariness and inconsistency.

The ideal type of bureaucratic organizations constitutes a top-down perspective, where strict implementation of rules and, hence, the control of bureaucrats’ actions is central. From this perspective, scholars have studied the functioning of bureaucracies. Within this perspective, scholars started to acknowledge the inevitability of the human factor in bureaucratic organizations, and studied the rationality of individual bureaucrats’ decision making (Downs, 1966; Simon, 1976). It was realized that bureaucrats’ cognitive capabilities to process information are limited, and that they use heuristics such as shortcuts or standard operating procedures to reduce uncertainty (Simon, 1976; Jones, 2001). Besides this ‘bounded rationality’, scholars also

acknowledged that bureaucrats' individual preferences and personal standards play a role in discretionary decision making (Lipsky, 1980). Within the top-down perspective on bureaucracy these idiosyncrasies pose a problem of control.

In line with this, scholars have studied bureaucrats' accountability towards the top (e.g. Bovens, 1998). Principal-agent theorists, likewise, have studied how lower order bureaucrats carry out requests of higher order principals (e.g. Brehm & Gates, 1997). Literature on policy implementation, in the same vein, focuses on the 'vertical dimension' of public government (Hupe et al., 2015, p. 11). Implementation studies typically look at whether and how bureaucrats' practices align with formulated policy goals (Brodkin, 2015). Such studies usually point to the gap between official policy and bureaucrats' decisions, which has been both positively and negatively framed (see Gofen, 2013 on a discussion of these framings). The common denominator of these studies is that they compare policy 'as prescribed' with the practice of policy implementation. Within these perspectives, bureaucrats' actions and decisions are perceived as sources of uncertainty, possibly entailing a loss of control at higher levels in the public organization.

Uncertainty reduction is thus central to the ideal type and top-down view of bureaucracy. Street-level bureaucracy scholars commonly focus on service delivery or law enforcement at the lowest level of government, and have acknowledged that some irreducible uncertainty is part and parcel of discretionary decision making (Lipsky, 1980; Maynard-Moody & Musheno, 2003). This dissertation has studied street-level officials' uncertainty experiences and particular ways by which they deal with uncertainty.

1.2. Uncertainty in street-level work

The street-level bureaucracy literature has focused on the bureaucrats at the frontline of government decision making and implementation, who have contact with citizen-clients and have considerable discretion in making decisions (e.g. Lipsky, 1980; Prottas, 1979). Street-level bureaucrats have been characterized as bureau-professionals, who exercise professional discretion combined with principles of rational administration (Newman, 1998). With the recognition that street-level bureaucrats have leeway in making decisions, the idiosyncrasies of frontline decision making became a relevant scholarly subject. Davis (1969, p. 4) defines discretion as follows: 'A public officer has discretion wherever the effective limits on his power leave him free to make a choice among possible courses of action and inaction'. Discretion of frontline officials is necessary, it is held, because policies, rules and laws are never specific enough to fit particular cases (Hoag, 2011), to be able 'to respond to the unexpected and to ensure that services are responsive to individual need' (Evans, 2015, p. 281). Discretion, as fundamentally a top-down notion (Maynard-Moody & Musheno, 2012) thus implies a degree of uncertainty in bureaucratic organizations, possibly opening the door to inconsistency and particularism.

In recent decades, street-level work has undergone transformations which allegedly hamper frontline workers' discretion. Firstly, under the heading of New Public Management, managerial reforms have subjected frontline workers' performances to closer scrutiny and evaluation (Brodkin, 2015). Whereas scholars reflecting upon the effects of managerial reforms on street-level discretion still perceive of street-level work as inherently discretionary, the common assumption is that this human judgment has become more standardized, that is, that street-level professionalism is constrained by managerialism (Bovens & Zouridis, 2002; Clarke & Newman, 1997). Secondly, the increasing use of ICT within bureaucratic organizations is believed to restrict street-level discretion and to even make it redundant (Bovens & Zouridis, 2002; Buffat, 2013). ICT enables automation of routine cases, thereby reducing managers' uncertainty (Schuppan,

2015; Snellen, 1998). Moreover, it is held, the discretionary decisions are no longer made by street-level bureaucrats, but by system developers and ICT experts (Bovens & Zouridis, 2002).

Some scholars, however, argue and show that street-level bureaucrats find ways to resist managerial constraints and still retain their discretionary powers (Brodkin, 2011; Soss et al., 2011a). Although street-level bureaucrats may be held accountable to their managers to a larger extent than before the managerial reforms, there is also a trend towards participatory and democratic forms of governance, making accountability more horizontal (Hupe & Hill, 2007). Horizontal relations between public officials and citizens, give the latter opportunities to participate in decision making (Bartels, 2013). Also, studies have shown that ICT is not simply a constraint to frontline discretion, but also extends discretion, by obscuring officials' informal use of discretion (Jorna & Wagenaar, 2007), by providing additional action resources (Buffat, 2013), and due to the subjectivity and interactivity of frontline work which cannot be handled by ICT (Schuppan, 2015). In fact, other studies have suggested that street-level bureaucrats' leeway in making decisions has become even larger (e.g. Dubois, 2014). The shift towards more horizontal service provision and law enforcement put frontline officials' perceptions and interpretations of citizen-clients' attitudes and behaviors upfront. Notions such as trust, responsiveness and collaboration are not predefined, but left open for frontline officials' professional expertise (Dubois, 2014; Evans, 2015). Whereas more discretion and hence uncertainty inevitably lead to diversification and individualization at the ground floor of government, it is also a means for the government to exert 'power over citizens' behaviors' (Dubois, 2014, p. 38), since it allows officials to enter and assess citizen-clients' private lives.

There is, however, no insight in the uncertainties bureaucrats themselves experience. Although the street-level bureaucracy literature acknowledges and shows that the practice of service delivery and law enforcement is not unambiguous and straightforward, and sometimes highly uncertain (e.g. Lipsky, 1980; Maynard-Moody & Musheno, 2003; Vinzant & Crothers, 1998; Wagenaar, 2004), there has been no scholarly attempt to systematically study the sources, kinds and responses to uncertainties as experienced by bureaucrats themselves. Whereas more 'professional' leeway entails more uncertainty at the top, possibly leading to inconsistency, there is no insight in how frontline bureaucrats themselves, in a context encouraging professional judgment, experience uncertainties and endeavor to reduce these uncertainties. This is an important topic of study, because it gives insight in the aspects of frontline work that are commonly not visible, though likely to affect the transparency and predictability of frontline service provision and law enforcement. Furthermore, bureaucratic organizations' emphasis on professional judgment and horizontal coordination likely leads to breaking up policy 'into local inter-individual arrangements' (Dubois, 2014, p. 39), making the social context of frontline decision making more important.

This dissertation, therefore, has employed a bottom-up perspective, allowing for the study of uncertainty from officials' own experiences. Contrary to studies assuming that uncertainty is part and parcel of discretionary decision making, this study has examined the uncertainties experienced by officials who have been granted considerable discretion. By reviewing the classical literature on bureaucracy and street-level bureaucracy, this dissertation has argued that a more sociological notion of uncertainty was largely missing. Using a sociological approach, this study allowed for studying officials as part of a larger bureaucratic context, including social interactions and relations. Based on an analysis of frontline officials' stories, this dissertation has pointed to the social character of frontline uncertainty.

1.3. Reducing uncertainty: Social network, signals and stereotypes

This dissertation, furthermore, has focused on three typical uncertainty reduction behaviors: deliberation with colleagues, the use of signals and cues, and the use of stereotypes. Street-level bureaucracy scholars typically focus on officials' coping strategies, when they are faced with limited resources, time and high case- and workloads (Lipsky, 1980; Tummers et al., 2015). These studies show how coping strategies such as creaming, routinizing and rationing are used to achieve a manageable workload. This current study goes beyond the coping literature, by not viewing officials' behavior as an expression of dealing with psychological stress or limited resources, but to look at how officials deal with uncertainty more generally.

The street-level bureaucracy literature suggests frontline officials rely on their fellow workers when faced with ambiguity (Keiser, 2010; Lipsky, 1980; Rutz et al., 2015). Frontline officials' interactions with peers and citizen-clients are nowadays even employed as government tools to increase responsiveness (e.g. Rutz et al., 2015). New forms of governance introduce social dynamics in the form of trust and collaboration between officials and citizens, and among peers as means to increase legitimacy and performance (e.g. Bruhn & Ekström, 2017). Furthermore, new organizational arrangements enabling collaboration between different agencies and professionals to deal with multiproblem clients and complex problems have arisen (Groeneveld & Van de Walle, 2011; Noordegraaf, 2011, 2016).

Notwithstanding this emphasis on interactions, studies on the impact of the frontline social context on decision making remain scarce. The same applies to research on frontline workers' use of simplifications and stereotypes to cope with limited resources and work pressure (e.g. Andersen & Guul, 2016; Lipsky, 1969; Mennerick, 1974); this line of research primarily relates bureaucratic working conditions to individual bureaucrats' use of shortcuts and stereotypes. Sociological explanations of the use of stereotypes by officials are less often studied within public administration research. Without disregarding the psychological aspects of stereotyping and the effect of bureaucratic working conditions on this, this dissertation examines a sociological explanation of stereotyping (e.g. Harrits & Møller, 2014; Soss et al., 2011b). This approach assumes that in categorizing others, people inevitably rely on broader social categories, and allows for the study of frontline stereotyping beyond coping. Using the sociological status characteristics and double standards theory (e.g. Ridgeway, 1991; Foschi, 2000), this study has explored and tested whether and how social categories could also serve as frames affecting the interpretation of signals.

Whereas the research on coping offers us valuable insights in frontline officials' behaviors and decision making, it seems to equate the use of discretion to officials' responses to stress resulting from the higher up demands (Evans, 2015). In line with Evans (2015), Gofen (2013) and Maynard-Moody and Musheno (2003; 2012) this dissertation has employed a bottom-up approach that allowed for the study of the broader social context of frontline behavior. Within the remainder of this chapter, the research aims and questions will be presented, the research approach and setting elaborated upon, the theoretical, methodological and societal relevance discussed and, lastly, the outline of the overall dissertation presented.

1.4. Research aim and general research question

This dissertation thus aims to understand the kinds and ways of dealing with uncertainties experienced in a bureaucratic context where frontline officials are encouraged to rely on their own standards to assess citizen-clients. This context has been the focus of this research, and has been a constant throughout the study (the specific research setting will be discussed in section 1.5).

The overall research question that is central to this dissertation is as follows:

What kind of uncertainties do frontline officials – working within a bureaucratic context that has expanded frontline leeway and encourages professional judgment – experience, and how do they attempt to reduce these uncertainties?

This general question will be answered in the conclusion and discussion chapter of this dissertation. The four empirical chapters address different sub-questions and together provide the answer to the overarching research question. Chapter two, the first empirical chapter, aimed to understand uncertainty at the frontline, and focused on the following research question: What kind of uncertainties do frontline tax officials experience, how do officials respond, and what are the sources of these uncertainties? Chapter three focused on the social context of frontline decision making, and aimed to answer the following research question: How do social dynamics within different frontline contexts impact case-related decision making? Chapter four explored whether officials evaluate similar signals differently for different social groups: How does citizen-clients' belonging to social groups affect officials' interpretation of signals? In order to answer this main research question, this study first examined the signals officials look at to assess citizen-clients' trustworthiness and the way they do this. This study allowed for examining officials' interpretations of signals and the role of stereotypes in this. Following up on the findings of chapter four, chapter five deductively tests whether officials use double standards in evaluating the trustworthiness of citizen-clients from different status groups. The following research question was central: Do citizen-clients' status characteristics affect frontline officials' evaluations of similar signals unequally, and if so, how?

1.5. Research approach and setting

In order to study frontline uncertainty and uncertainty reduction from a bottom-up perspective, a research approach was chosen that allowed to contextualize frontline officials' work. The methodology and methods chosen enabled me to study the experiences, interpretations and meaning making of frontline officials. In what follows, the methodological presuppositions underlying this dissertation and the methods used will first be discussed. Secondly, the research setting of this dissertation will be explicated. Subsequently, tax officials will be positioned as particular types of street-level bureaucrats.

Methodology and methods

This dissertation's main focus is on the experiences, judgments and decision making of frontline public officials, and is particularly interested in uncertainty experiences, the social context of decision making, and the role of stereotypes in officials' evaluations of citizen-clients. Within this dissertation, I assume these objects of study are socially constructed. In studying these objects, this dissertation not only aims to understand officials' own meaning making, but also to explain their evaluations by looking at more structural factors. This study thereby not only assumes that people's agency has causal power, that is, people who have intentions and reasons to act in certain way, but that social structures, such as social class or educational attainment, can also exert an influence on people's interpretations and behavior (e.g. Alexander, 2005; Harrits, 2011). These social structures have an 'objective force' and need not to be understood by people themselves in order to exert themselves (Harrits, 2011). As such, this dissertation is interested in understanding officials' lifeworld, consisting of people's inner and intersubjective

meanings and practices, but also in explaining their behavior from an ‘observer’s perspective’ (Harrits, 2011, p. 160).

The methodological approach, thus, is aimed at generating insight in officials’ own meaning making and at generating knowledge about social structures that are the conditions of social action. This dissertation has used interpretive research methods, such as storytelling, participant observation and in-depth interviews to study officials’ experiences, practices and interpretations. This part has given insight in people’s own accounts and explanations of their actions. To study the more implicit influence of social stereotypes on officials’ evaluations, this dissertation has conducted a policy-capturing study. This design has allowed me to test the sociological double standards mechanism, using quantitative analysis methods.

Dutch tax authorities: Responsiveness and professional judgment

This research has primarily been conducted at the Dutch tax administration, which is a public organization that operates under the Ministry of Finance and has the responsibility to levy and collect taxes and social insurance premiums. It furthermore has the task to detect tax fraud and to monitor compliance with tax laws. The organization largely works according to the ideal type of rational-legal bureaucracy, where uncertainty reduction is pursued by standardization of work processes and hierarchical control. In the division of small and medium sized enterprises, which is the research setting of this dissertation, standardization is strived for at different levels. Firstly, at the system level, statistical models are used and applied to assess entrepreneurs’ tax compliance (Leviner, 2009). This means that pre-established models determine the kind of citizens that are being considered ‘suspicious’ and selected for further investigation. Besides that, audits are conducted among randomly selected enterprises, of which the outcomes are used to inform risk group policy and automatization processes (Belastingdienst, 2016).

Secondly, frontline tax officials, who are endowed with the task to carry out the audit on-site, are encouraged to extensively prepare their audit in the ‘preplanning phase’ before they go on-site (Belastingdienst, 2016). Aided by digital systems containing information about the to-be-audited enterprise, officials need to make calculations within the digital system, which in turn gives a general direction with regards to the auditing approach, that is, the points for attention. The final decisions also have to be processed within the digital system. Although tax officials still have much leeway in deciding upon their approach (how they are going to audit an enterprise), the preparation phase ensures officials inspect enterprises well informed, as to increase uniformity and to minimize the burden put on enterprises. The tax administration furthermore employs team managers and audit managers, who serve as an additional check on the decisions made by frontline tax officials. Team managers focus more generally on the achievement of organizational goals, which involves both quantity and quality of inspections, and the ‘throughput time’ of these inspections. The audit managers allocate cases to tax officials based on their expertise, and review tax officials’ final reports.

Whereas tax officials are, thus, typical bureaucrats who have to give account to their managers, and are guided by hierarchical rules, they have considerable discretion in how they conduct their inspections and in making decisions. It is their task to determine whether something is wrong, that is, whether there are gaps within the bookkeeping records or other violations of fiscal laws. Frontline tax officials have to assess whether possible gaps came about by mistake, negligence or fraud, and to decide about the concomitant correction and possible fine. This not only involves expert, fiscal knowledge and skills, but also more practical knowledge needed to assess entrepreneurs’ stories, competences and intentions.

These practical forms of knowledge have even come to play a more prominent role in front-

line practice, where officials work with the 'horizontal supervision' approach since 2005. This policy has entailed a shift from a vertical command and control approach to responsive and collaborative regulation and enforcement (Gribnau, 2007). Trust between officials and entrepreneurs is regarded as a means to increase the latter's compliance with regulations. As a consequence, the interactional processes between officials and taxpayers are seen as crucial in fostering compliant behavior. This horizontal policy encourages officials to assess tax returns on their acceptability, rather than their mere correctness. This means that officials are encouraged to collect 'sufficient' information to make a judgment, and have 'to do enough work, but not too much' (Belastingdienst, 2016, p. 4). In practice, this means officials are discouraged to look for mistakes, and to accept tax returns when there is sufficient evidence to do so. This involves a large degree of professionalism or 'professional judgment' (Belastingdienst, 2016, p. 5). Tax officials are moreover allowed to make settlement agreements with entrepreneurs¹, when there is disagreement about interpretations of ambiguous rules and legislation. This shift towards horizontal regulation offers frontline tax officials more leeway in determining what is acceptable, and hence to make judgments according to their own criteria.

Professional frontline teams

Chapter two combines data collected within the Dutch tax administration and data gathered within the Belgian Inspectorate of Social Laws. The latter have been collected as part of another doctoral dissertation (Loyens, 2012). Combining data from different cases has allowed us to study the impact of social dynamics on frontline decision making in different social constellations – different types of frontline teams and one-on-one interactions with citizens. Within the Dutch tax administration, some officials are part of the multidisciplinary 'take-away team' aimed at combating organized crime and confiscating criminal assets. It involves a collaboration between five large municipalities in the province of Brabant in the Netherlands, the public prosecutor, the police, Fiscal Information and Investigation (FIOD), the tax authorities, and the Royal Military Police. The different parties exchange information about suspects, and collaboratively make a decision.

The officials within the Belgian inspectorate inspect undeclared work violations and enforce wage and labor condition regulations, during on-site visits. Just as tax officials, these labor inspectors have considerable discretion; they can decide to make a report, propose a settlement agreement or give a warning. Within this inspectorate, social interactions are emphasized too, in the form of interorganizational frontline teams. Anti-fraud teams with labor inspectors from the five Belgian labor inspectorates focus on undeclared violations, and have been set up to increase consistency in decision making across organizations. The interdisciplinary team, furthermore, involves representatives of the police, tax administration, food inspectorate and labor inspectorate, who all focus on their own specialty, and which is aimed at enhancing efficiency and decreasing the frequency of inspections. The Dutch 'take away' team involves shared discretion, that is, leeway given to the team to make decisions, whereas the Belgian teams involve separate decision making tasks, or one decision making task with one responsible inspector. By combining the different data sources, the social dynamics within different constellations could be analyzed.

1. The words entrepreneur(s) and citizen-client(s) are used interchangeably in this dissertation.

Tax officials as street-level bureaucrats

Tax officials can be seen as a particular type of street-level bureaucrats. In accordance with the street-level bureaucrats described by Lipsky (1980), tax officials have face-to-face encounters with citizen-clients and have considerable discretion in decision making (Cohen & Gershgoren, 2016; Fineman, 1998; Nielsen, 2015). Their work is characterized by tax laws and regulations that also leave much leeway for interpretation. They also have to work under typical bureaucratic conditions, such as time pressure, high workload and incomplete information (Cohen & Gershgoren, 2016). As such, they are similar to other street-level bureaucrats, such as social workers and police officers.

However, there are certain characteristics of tax officials' work which make them different from other street-level bureaucrats. First, tax officials work within a fairly rationalized bureaucratic environment, their work can be quite complex requiring expert knowledge (Nielsen, 2015), and tax officials possess a lot of prior information about the inspected citizen-client (Cohen & Gershgoren, 2016). Second, interactions are typically initiated by tax officials who are mandated to inspect enterprises, and entrepreneurs are required by law to disclose information regarding their bookkeeping records. This means that entrepreneurs cannot simply exit these interactions (Nielsen, 2015). These characteristics make tax officials a powerful type of street-level bureaucrats, compared to other street-level workers whose work is characterized by less discretion, less expert knowledge and where interactions with citizen-clients are initiated by citizen-clients themselves. The focus on this particular type of street-level bureaucrats has enabled the study of uncertainty and uncertainty reduction in a fairly rationalized bureaucratic context.

1.6. Theoretical and methodological relevance

This dissertation has employed a sociological, bottom-up approach to studying frontline work, judgment and decision making. Looking at the impact of *social context* of frontline work and *broader social stereotypes* on frontline decision making, this dissertation makes a substantive contribution to the street-level bureaucracy literature. Moreover the bottom-up approach involves a particular choice for research methods, aimed at understanding and explaining street-level work and judgment, *without losing sight of context*. The sociological, bottom-up approach, furthermore, has implications for the literature on new models of governance which perceive of encounters between officials and citizens as valuable tools to increase legitimacy, responsiveness, and compliance.

Sociological approach to the study of street-level work and stereotyping

The street-level bureaucracy literature typically focuses on explaining street-level bureaucrats' use of discretion, by looking at the micro-level interactions between individual frontline officials and citizens (e.g. Maynard-Moody & Musheno, 2003), the influence of either characteristics of the official or the citizen-client on decision making (e.g. Harrits & Møller, 2014), or by looking at the influence of macro-level factors such as political and managerial influences (e.g. May & Winter, 2009). However, the social context of the decision making process is often left out of the equation. The street-level bureaucracy literature suggests frontline officials rely on their fellow workers when faced with ambiguity or role conflict (Keiser, 2010; Lipsky, 1980; Rutz et al., 2015). Officials not only deliberately rely on peers for advice and support (e.g. Maroulis, 2015), but also implicitly take into consideration what peers would do in similar situations when they have to make decisions (Keiser, 2010). There is however no insight in

how social dynamics between public officials, but also in official-citizen interaction could affect decision making about citizens. By scrutinizing social dynamics in decision making about citizen-clients, this dissertation takes into account the meso-level of street-level work.

There has been much scholarly attention to the study of discrimination by street-level officials. Lipsky (1980) perceives it as a way of coping with the limited resources and time in bureaucratic contexts. Within public administration research, there is a nascent trend to study street-level discrimination using experimental research designs (e.g. Andersen & Guul, 2016) and to study frontline behavior relying on psychological theories (e.g. Grimmelikhuijsen et al., 2017; Tummers, 2016). Although these studies offer valuable insights in stereotyping and behavior at the frontlines, they typically assume individualistic and sometimes psychological explanations of discrimination (see also Gambetta & Hamill, 2005), and do not consider or look at sociological explanations of stereotyping. In line with more sociological studies on the use of stereotypes by public officials (Dubois, 2014; Epp et al., 2014; Harrits & Møller, 2014; Schram et al., 2009), this dissertation studies whether and how stereotypes persistent in society affect frontline judgment, by exploring and testing a sociological mechanism of stereotyping. This thesis uses the sociological status characteristics and double standards theory to examine if and how frontline officials use double standards in evaluating citizen-clients from different status groups. The theory proposes that stereotypes serve as frames affecting the interpretation of similar evidence unequally, that is, that the evaluation is stricter for the lower status group than for the higher status group (Ridgeway, 1991; Foschi, 2000). This dissertation adds to the public administration literature by going beyond the study of stereotypes as informational shortcuts, to exploring and testing whether and how stereotypes could also serve as frames or 'moderating contexts' unequally affecting the evaluation of similar signals.

Methods to study social dynamics and implicit stereotyping in context

A variety of methods have been used to study street-level bureaucrats' work and decision making. A disciplinary distinction is visible with regard to the methodologies and methods that have been used. Sociological studies typically use interpretive methods, such as storytelling, in-depth interviewing and participant observation to study street-level bureaucrats' categorizations, judgments and experiences (e.g. Dubois, 2013, 2014; Maynard-Moody & Musheno, 2003, 2012). These studies show how categorizations of citizen-clients are made, aided by social stereotypes. Public administration scholars have started studying frontline behavior and discrimination using experimental research designs using treatment and control groups (e.g. Andersen & Guul, 2016). Whereas the first branch of literature offers us rich descriptions of frontline reasoning, these qualitative designs do generally not allow for studying effects of certain stereotypes or officials' background characteristics on frontline judgment. The second branch of literature, to the contrary, does allow for the study of such effects, but the findings of these studies are generally less ecologically valid, that is, more difficult to generalize to real-life settings.

The bottom-up approach of this dissertation has motivated a choice for the use of specific research methods that allow for the study of context. Firstly, this study used inductive research methods, such as storytelling and participant observation, to study frontline officials' meaningful experiences, judgments and decision making in their broader social context. The storytelling method allowed for gaining insight in officials' work from their own perspective, focusing on how officials themselves make sense of the uncertainties in their day-to-day work. Participant observation, moreover, enables researchers to study officials' behavior and meaning making in their actual work contexts. These methods, furthermore, were suitable for studying the role of

social dynamics and interactions which were implicit in officials' stories, explicitly mentioned by the respondent, or observed by the researcher.

Secondly, based on the findings of the qualitative studies, this dissertation has used a policy-capturing design that allows for the study of effects of stereotypes on frontline judgment, while not losing sight of context. It entails letting respondents evaluate multiple hypothetical, but realistic scenarios within an interview setting. This practice resembles real-life settings where officials evaluate cases in comparison to other cases. Moreover, with the help of the in-depth stories told by frontline tax officials, the scenarios were constructed as to resemble real-life settings. As such, the method allowed for the study of effects, and approximates the street-level context better than traditional experiments.

The centrality of bureaucratic encounters to contemporary governance

The relationship between frontline officers and citizen-clients has always been at the core of street-level bureaucracy literature (Bartels, 2013). It has gained importance with the introduction of new public management reforms that have led to the reformulation of the bureaucratic encounter as a businesslike arena, where officials are responsive to citizen-clients' needs and demands (Vigoda, 2002). With the aim to increase public organizations' performance and legitimacy, citizens have become more involved in the process of public service delivery. Post-NPM movements, based on notions of participatory governance, even point to a more horizontal relationship, where officials collaborate with citizens and other stakeholders in service delivery networks to deal with complex problems (Bartels, 2013; Vigoda, 2002). The process and outcome of frontline decision making has hence come to depend on interactions between officials and citizens (Bartels, 2014; Rice, 2013). Trends towards more responsiveness are also visible in public organizations preoccupied with law enforcement such as tax and inspection authorities (e.g. Mascini & Van Wijk, 2009).

This has several consequences: social interactions are brought to the center of frontline decision making (Bartels, 2013; Bruhn & Ekström, 2017) and it has led to an individualization of service provision and law enforcement, giving frontline workers more leeway to decide what is, for instance, responsiveness in specific cases (Dubois, 2014; Rice, 2013). While more discretion offers bureaucrats more possibilities, it is also likely that frontline officials experience uncertainty in decision making. This dissertation has studied the kinds, sources and responses to uncertainties experienced by bureaucrats working in such a context. It thereby gives insight in the uncertainties that are inherent to discretionary work and likely affect bureaucratic decision making, but which have remained underexplored. Also, despite the fact that new forms of governance introduce social dynamics in the form of trust, collaboration and responsiveness between officials and citizens, and among peers as tools to increase legitimacy and performance, the study of the impact of social context and interaction on decision making lags behind (Bruhn & Ekström, 2017). This dissertation has made an effort to address this gap and studied how social dynamics may impact on decision making about citizen-clients.

The emphasis on tailor-made services and responsiveness in bureaucratic organizations, furthermore, shifts focus from following strict rules to making decisions based on professional judgment. Notions such as responsiveness, trust and activation, which play an increasingly prominent role in frontline decision making, are not predefined, but left open to frontline professionals to flesh out in practice (Dubois, 2014; Evans, 2015; Rice, 2013). In other words, frontline officials are encouraged to use their own criteria in evaluating and judging about cases, in order to be more responsive towards particular cases. This makes the study of officials' own interpretations a relevant scholarly topic. Although undefined and vague notions as trust and

risks are criteria governing the frontline practice of law enforcement or service provision, there is no insight in how frontline bureaucrats evaluate whether someone is trustworthy or risky or not. Such characteristics are not readily observable and need interpretation work. This dissertation has studied what signals frontline officials look at to categorize citizen-clients as either trustworthy or untrustworthy, how they do this, and the role of stereotypes in these evaluations.

1.7. Societal relevance

The findings of this dissertation also have societal relevance. Societal and economic changes, such as the aging population, and the financial crisis and concomitant increase in unemployment, have put pressure on welfare states to cut back on government spending and/or to increase taxes (Clayton & Pontusson, 1998; Vis et al., 2011). These societal changes also impact the role and tasks of street-level service providers who, on a daily basis, have to deal with growing client demands, but with diminished resources to provide services (e.g. Savi, 2014). Although new public management reforms have been said to put limits to street-level officials' discretion and ability to be responsive, there is also a shift towards horizontal steering arrangements, giving officials more room to make professional judgments and decisions (e.g. Van der Aa & Van Berkel, 2015). In fact, 'decentralized' solutions are sometimes deployed to deal with crisis at the level of service delivery, that is, difficult decisions regarding how to allocate scarce resources are delegated to the frontline (Savi, 2014). These developments have contributed to a more positive view on frontline discretion, 'as an asset for creative, deliberative, and informed judgment' (Bartels, 2013, p. 473). Within participatory forms of governance, discretion is even valued over strict rule following and impartial treatment, since it is believed to enable deep collaboration between officials and other parties (Bartels, 2013).

Nowadays, discretion thus has acquired a positive connotation, not only to public agencies and officials who promote 'professional judgment', but also by scholars who started studying its role in deliberative forms of governance. Nevertheless, it remains important for citizens, public officials and politics, to study uncertainty and uncertainty reduction in public organizations. Giving frontline officials considerable leeway to make decisions involves a more indeterminate bureaucratic organization. Studying how officials in such a context experience uncertainty, their responses to it, and their ways of managing it, therefore, is all the more important, since this likely has implications for the way they treat and make decisions regarding citizens, and the way in which they fulfill their role as 'policy implementers'.

The findings of this dissertation are of relevance for several societal actors. Firstly, the study gives insight in the uncertainties experienced by frontline workers in bureaucratic context where they are encouraged to rely on their own judgment to assess citizen-clients. This thesis also gives insight in whether and how these officials use stereotypes and double standards in evaluating citizen-clients. As such, this thesis offers valuable insights to policy makers, managers and frontline workers themselves. Secondly, the findings of this thesis are of relevance for citizens too. At the other side of the interaction, they are the 'subject' of frontline decision making. Although citizen-clients are increasingly made part of decision making processes to increase for instance collaboration, trust, or compliance, there are still aspects of decision making that remain implicit and thereby invisible for citizens, such as uncertainties and stereotyping. Knowledge about such covert and implicit practices could contribute to a more knowledgeable and hence more powerful position of citizens in encounters with officials, who may challenge officials' categorizations and find ways to influence frontline decision making.

1.8. Structure and outline of dissertation

The overarching research question has been broken down to four sub-questions that are central to the four empirical chapters of this dissertation. The research questions posed and the methods used in each chapter build on the findings of prior chapters. Chapter two addresses the kinds, sources and responses to uncertainties experienced by frontline tax officials working with more leeway. By using an inductive research approach and the storytelling method, this study has found three notions of uncertainty that are at play: information, interpretation and action uncertainty. Structural aspects of frontline work, external factors such as economic hardship, and the 'horizontal supervision policy' have been found to underpin these uncertainties. The study has also suggested that officials typically respond by trying to find more proof, relying on their colleagues to reach a supported account, and by improvising on-the-spot in interaction with citizen-clients.

Chapter three builds on the finding of chapter two indicating that officials rely on colleagues in case of doubt, by specifically focusing on the social context of frontline decision making. The third chapter more specifically explored how social dynamics in officials' interactions with colleagues and citizens may impact decision making about citizen-clients. By combining inductive, qualitative data of the Dutch tax administration and the Belgian labor inspectorate, this study did not only focus on officials' deliberate reliance on colleagues and peers, but also on the more implicit social dynamics between officials and officials and citizens. The findings of the study indicate that social dynamics could either lead to uncertainty reduction for the individual bureaucrat, but also lead to arbitrariness in decision making, and thus, bureaucratic uncertainty.

Chapter four explores the validity of the double standards mechanism by looking at how similar signals may be interpreted differently for different social groups. In order to answer this question, this study has first examined what signals officials look at to categorize citizen-clients as either trustworthy or untrustworthy, and how they come to interpret certain attributes as signals. It builds on chapter two's findings suggesting that officials' image construction of citizen-clients and their enterprises is based on other factors than sole information. In fact, that study has shown that proving and substantiating one's own account is part and parcel of frontline work. It also builds on chapter three's findings which point to the subjectivism of frontline decision making. This suggests that officials may use signals and stereotypes in assessing citizen-clients (e.g. Gambetta & Hamill, 2005).

Based on in-depth interview data, the study has shown that officials not only look at signals related to entrepreneurs' bookkeeping, but also their demeanor, home situations and belonging to social groups to assess their competences and intentions. Relying on the sociological status characteristics and double standards theory, this study *explores* how signals from lower status groups in society, such as people from lower social classes, may be evaluated stricter (more negative) than similar signals from higher status groups, such as people from higher social classes. The findings have suggested that similar signals may be interpreted differently for different social groups. However, this was only found for a minority of respondents, and findings point in different directions.

In order to more systematically *test* the sociological double standards propositions, chapter five employed a policy-capturing study. This chapter builds on the findings of the previous chapter. Chapter four explored the possibility of the use of double standards by looking at officials' interpretations. It used an *inductive* approach and revealed how interpretations of signals may differ across different social groups. Relying on the specific findings of this study, chapter five more deductively tested whether officials evaluate similar signals stricter for lower

status citizen-clients, than for higher status citizen-clients. The design of the study of chapter five allowed for more control of the research context. It involved letting respondents evaluate a fairly large number of scenarios, portraying inspections at enterprises, where cues indicating the quality of bookkeeping, quality of interaction and status characteristics were manipulated. Using multilevel analyses, this study has found some signals are indeed evaluated differently for lower status entrepreneurs than higher status entrepreneurs, but most hypotheses were not supported. The disadvantages and advantages of the policy-capturing approach to study stereotyping have been discussed in comparison with other methods such as interviewing and traditional experiments.

Chapter six concludes by answering the general research question, by discussing the specific theoretical, methodological and societal contributions, the limitations of this dissertation and recommendations for future research.

Chapter 2

How to prove, how to interpret and what to do? Uncertainty experiences of street-level tax officials

Abstract

This study examines the kind of uncertainties frontline tax officials working with a trust-based inspection approach experience in interacting with citizen-clients. The classical literature on bureaucracy and the street-level bureaucracy literature suggest frontline officials face two kinds of uncertainties: information and interpretation problems. Analyzing stories of Dutch frontline tax officials collected through in-depth interviews, this article shows that these two kinds of uncertainty only explain a part of the uncertainties experienced. Respondents also face action problems requiring improvisational judgments. The study furthermore finds that different sources underlie these uncertainties, pointing to possible explanations.

Raaphorst, N. (2017). How to prove, how to interpret and what to do? Uncertainty experiences of street-level tax officials. *Public Management Review*. Doi 10.1080/14719037.2017.1299199

2.1. Introduction

Street-level bureaucrats' interactions with citizens have been studied extensively within the domain of public service provision (e.g. Goodsell, 1981; Katz et al., 1975; Lipsky, 1980; Prottas, 1979). The street-level bureaucracy literature mainly focuses on how frontline officials use their discretion, that is, the room for maneuver they have within the given rules and regulations (e.g. Harrits & Møller, 2014; Lipsky, 1980). Studies within this branch of literature often point to the open-endedness, ambiguity and uncertainty of everyday administrative work (e.g. Dubois, 2014; Hoag, 2011; Maynard-Moody & Musheno, 2003; Wagenaar, 2004). In fact, uncertainty is inextricably linked to discretionary decision making; bureaucrats 'derive their institutional necessity from these uncertainties' (Dubois, 2014, p. 41). Although some scholars reflect on this frontline uncertainty (e.g. Dubois, 2014; Fassin, 2013), there is still lack of understanding of the kinds, conditions and consequences of uncertainty at play in frontline work.

A deeper understanding of how officials respond to uncertainties, is especially important in light of managerial reforms that have put more emphasis on trust and collaboration between officials and citizens. New Public Management reforms were aimed at enlarging citizen-clients' choice in public goods, but also at improving public officials' performance and responsiveness (Bartels, 2013; Pollitt & Bouckaert, 2004). Governance styles aimed at involving citizens, organizations and other stakeholders in the process of policy-making and implementation embrace notions as trust and collaboration. This is not only true for social welfare agencies, but also for organizations engaged in the more traditional regulation and law enforcement functions of the government, such as inspection agencies and tax authorities (e.g. Leviner, 2009; Mascini & Van Wijk, 2009). Such interpersonal notions as trust and collaboration are not predefined, but assumed to be part of officials' professional expertise (Evans, 2015).

In fact, horizontal interactions between officials and citizens aimed at collaboration and trust give frontline officials more leeway to act on the situation at hand, which makes the bureaucratic process less determined (Dubois, 2014). If bureaucrats' actions are increasingly made dependent on their perceptions of citizens in interactions, and to a lesser extent prescribed by formal rules, this leads to a more uncertain bureaucratic process. This, then, spurs the question what kind of uncertainties street-level bureaucrats experience in their day-to-day task and how they navigate these. This is an important topic of study, because it gives insight in the aspects of frontline work that are commonly not visible and likely affect the transparency and predictability of frontline service provision and law enforcement.

This study aims to theoretically advance our understanding of frontline officials' experiences of uncertainty, within a policy context that emphasizes horizontal relationships with citizens, and where frontline officials are endowed with much leeway to make decisions based on their own interpretations. By studying Dutch frontline tax officials working with a policy that promotes responsiveness and trust, this paper aims to answer the following research question: what kind of uncertainties do frontline tax officials experience, how do officials respond, and what are the sources of these uncertainties? By focusing on tax officials, this study furthermore adds to the literature on street-level bureaucracy where the study of regulatory interactions is scarce (but see e.g. Cohen & Gershgoren, 2016; Nielsen, 2007, 2015)².

This study builds on different scholarly traditions on bureaucracy. By assessing the classical

2. The literature on regulation mainly focuses on conceptualizing and measuring regulatees' compliance to regulations, rather than on inspectors' use of discretion. Within this study, we are interested in how inspectors experience the uncertainty which is part and parcel of their discretion. For this reason, we mainly draw on the literature traditions that more broadly focus on frontline officials' use of discretion.

literature on bureaucracy and the street-level bureaucracy literature, this paper distils different notions of uncertainty. This study seeks to uncover whether (some of) these notions indeed describe the uncertainties experienced by frontline tax officials, or whether we need to broaden our understanding of uncertainties at the frontline. It furthermore explores the sources of these uncertainties, to gain insight in possible explanations of uncertainties at the frontline. After discussing the relevant literature traditions, the case selection and methods will be described, and the findings presented. This paper concludes with a discussion and directions for future research.

2.2. Review of literature

Uncertainty as information problem

The traditional model of bureaucracy has been, from its very beginning, concerned with uncertainty reduction within bureaucratic organizations (Weber, 1922/1947; Thompson, 1967/2003). Within this model, bureaucracies are seen as rational organizations that should limit individual bureaucrats' discretionary powers by setting strict rules and procedures. Technocratic knowledge, embodied in rules, procedures and policies, is put at the heart of bureaucratic organizations. The assumption is that rules and procedures could be applied directly to specific cases, without interference of the human factor, that is, individual bureaucrats' own interpretations (Thompson, 1967/2003).

With the recognition of the primacy of the human factor in bureaucratic organizations, theories of bureaucratic decision making also became a relevant scholarly subject. Simon (1976) argued that objective rationality as depicted by the ideal model did not reflect organizational reality. Organizational members' actual behavior is limited by incomplete knowledge about future consequences and by people's incapability of assessing all possible alternative behaviors (Simon, 1976; Downs, 1966). In this situation of bounded rationality, bureaucrats employ satisficing strategies in making decisions, and rely on relatively simple heuristics, shortcuts or standard operating procedures (Simon, 1976; Jones, 2001). Despite the limitedness of their cognitive capabilities, bureaucrats are still seen as actors that pursue goals by applying knowledge in order to resolve uncertainties prior to decision making. According to this perspective, uncertainty in the form of 'unknowns' is involved at all times. From a perspective that views technical knowledge as a means to control administrative work, 'unknowns' are deemed problematic. Uncertainty is, in this perspective, perceived as an information problem. Technical knowledge, albeit embodied in rules or standard procedures, is supposed to offer bureaucrats a sense of certainty, since having knowledge about a situation offers the possibility to master a situation and to act on it.

Uncertainty as interpretation problem

Since Lipsky's seminal work (1980) on street-level bureaucracy, the scholarly focus has shifted from a view on bureaucrats as 'cogs in machine' serving the larger powers, to a view in which bureaucrats' discretion in their daily encounters with citizens is acknowledged and considered a valuable subject of research (Bartels, 2013). Research has pointed out that bureaucrats' discretionary practices are not only informed by organizational classification systems and rules, but also by personal judgments regarding clients' worthiness or deservingness, based on cultural schemes, moral beliefs and values, or certain stereotypes (Dubois, 2013; Harrits & Møller, 2014; Lipsky, 1980; Maynard-Moody & Musheno, 2003; Mennerick, 1974; Prottas, 1979).

Within the literature on bureaucratic encounters it is argued, in this respect, that ‘there are no unambiguous criteria to discern citizen-client worthiness’, and that it is never ‘simple or straightforward’ (Maynard-Moody & Musheno, 2003, p. 119).

The street-level bureaucracy literature, then, sheds a different light on bureaucratic decision making. Knowledge is not simply ‘out there’ for bureaucrats to rely upon and apply in practice, but the ‘instances’ facing the organization need to be interpreted ‘in order to make these instances intelligible in the light of the organizational life-world’ (Handelman, 1978, p. 15). In creating a single, consistent and authoritative account of ‘what happened’, bureaucrats find themselves mulling over new information casting doubt on clients’ stories, leading them to re-interpret old information. Bureaucrats construct hypothetical story-lines that are only finalized by bureaucratic closure (Fassin, 2013; Handelman, 1978). The source of uncertainties, then, is not so much an absence of information, but rather a problem of interpretation of what ‘is really happening’.

Uncertainty of social interactions

Inherent to discretion is the assumption that it is only in the encounter with specific cases, that rules are enacted (Lipsky, 1980). Discretion at the frontlines ‘is necessary to respond to the unexpected and to ensure that services are responsive to individual need’ (Evans, 2015, p. 281). This emphasizes a more fundamental notion of uncertainty: the uncertainty that is always part of social interactions. The street-level bureaucracy literature points to the unpredictability of street-level work (Lipsky, 1980; Vinzant & Crothers, 1998; Wagenaar, 2004), and the tensions of face-to-face encounters (Dubois, 2014). Discretion always involves interaction between known abstract rules and yet unknown particulars. How these particulars manifest themselves is dependent on unpredictable interactions with citizen-clients.

This kind of uncertainty has received far less attention within the public administration literature on frontline work. The uncertainty that is inherent to discretion is treated as given. Insight in the uncertainties experienced by street-level bureaucrats themselves could give insight in the aspects of frontline work which are invisible, but which likely affect street-level behavior and decision making.

2.3. Research site

This research is conducted within the Dutch tax administration. It focuses on frontline officials who audit tax returns of small and medium sized businesses, and have face-to-face interactions with entrepreneurs as part of their job. Under the heading of the so-called ‘horizontal supervision’ approach, the Dutch tax administration has moved from a vertical command and control approach to responsive and collaborative regulation and enforcement (Gribnau, 2007). As a consequence, the interactional processes between officials and taxpayers are seen as crucial in fostering compliant behavior. This horizontal policy encourages officials to assess tax returns on their acceptability, rather than their mere correctness, which means officials are discouraged to correct little mistakes. Moreover, officials are encouraged to make settlement agreements with entrepreneurs when mistakes are encountered.

To assess the acceptability of tax returns, officials audit entrepreneurs’ bookkeeping records and operational processes. They also assess entrepreneurs’ intentions in order to advise the specialist who decides about the possible fine. Tax officials finalize their audit by writing a report in which decisions on corrections and possible fines are substantiated. When finished, this report is sent to the audit manager for a last check, before it is sent to the client. The cases tax

officials audit are selected on a central level in the organization, based on predetermined risks. It is officials' task to determine whether something is really wrong and whether it came about by mistake, negligence or fraud.

Like other street-level bureaucrats, such as social workers and police officers, tax officials have face-to-face encounters with citizen-clients and have considerable discretion in interpreting cases and making decisions (Cohen & Gershgoren, 2016; Fineman, 1998; Nielsen, 2015). Tax officials' work is characterized by a considerable amount of tax laws and regulations, but also by much leeway for interpretation. Tax officials moreover work under 'countless pressures and constraints' (Cohen & Gershgoren, 2016, p. 269), such as time pressure, incomplete information, ambiguous rules and regulation and conflicting work principles (e.g. Schott et al., 2016). In these respects, tax officials share the core characteristics of street-level bureaucrats. However, there are some characteristics of tax officials' work that set them apart from typical street-level bureaucrats. First, the rules and legislation tax officials work with can be complex and are not common knowledge (Nielsen, 2015), and officials possess much prior information about the inspected citizen-client (Cohen & Gershgoren, 2016). Second, interactions are typically initiated by tax officials who are mandated to inspect specific enterprises, and entrepreneurs cannot simply exit these interactions (Nielsen, 2015). These characteristics make tax officials a powerful type of street-level bureaucrats.

2.4. Methods

Seventeen frontline tax officials were interviewed, of whom fourteen are male and three are female³. Respondents were selected on theoretical grounds: respondents are all working with the horizontal supervision policy and have face-to-face interactions with entrepreneurs. This selection made it possible to study the uncertainties in a frontline context where rules and regulation have become less strict, and officials have more room for interpretation. Within this selection frame, this study aimed for a sample consisting of both male and female officials, as well as newcomers and veterans within the organization, as to be able to grasp a variety of experiences. Tax officials from different tax offices across the Netherlands were interviewed. The respondents were introduced to the study either by the researcher, or by their manager who approached them with the request for participation. The small number of women in the sample is due to the paucity of women working for the Dutch tax administration, and as such, corresponds with the male-female ratio in the organization as a whole. Ten respondents have been in service for over thirty years, one for eighteen years, and six have been in service for less than ten years.

This study focused on bureaucrats' detailed stories about face-to-face encounters with citizen-clients. Stories give insight in the meanings people attach to situations, and how they make decisions based on their own perceptions of situations (e.g. Maynard-Moody & Musheno,

3. To ensure ethical standards for data collection, different steps were taken. First, access to the organization was approved in advance by the director of the Dutch tax authorities and the director of the Small and Medium sized Enterprises (SME) division. As part of this, a confidentiality agreement was signed by the researcher, declaring that sensitive information about clients is treated as confidential. Second, respondents voluntarily participated to this research, and no consequences were attached to non-participation. Third, during the interviews, the interviewer ensured the respondents that the data (audio and transcripts) were processed anonymously, treated confidentially, and were not distributed to their managers or other people within the organization. The interviewer assured respondents there were no good or wrong answers and that they could quit the interview or ask questions at all times. Fourth, interview transcripts were sent to the respondents afterwards, which gave them the possibility to read the conversation.

2003). This method is preferable over direct questioning, since the latter method commonly yields rationalized answers, whereas storytelling probes respondents' meaningful experiences. Within an interview setting, respondents were asked to tell stories about situations they experienced as difficult or complicated. Within the first interviews the suitability of different interview questions to answer the research question was assessed. The author started out with questioning whether respondents could tell a story about an inspection, where they had face-to-face contact with citizen-clients, in which they experienced uncertainty. This did not yield rich stories, but mostly questions about what was meant by uncertainty. For this reason, uncertainty was replaced by experiences of difficulty or complexity. This phrasing turned out to respond to tax officials' lifeworld, since it yielded rich and long stories. Besides that, questions were asked about officials' daily activities, experiences at work, doubts, and interactions with entrepreneurs and colleagues. The interviews were audio-recorded and transcribed verbatim⁴.

Thirty-seven stories were selected for analysis. The situations described within the stories were all problematized, and all stories refer to experienced unknowns or ignorance. Within the analysis of the stories, the three uncertainties described above served as sensitizing concepts. Within the first step, the stories were coded using detailed codes that were still close to the data. Officials' responses were coded as well. The detailed codes were then confronted with the sensitizing concepts, in order to look whether they could be understood as a particular kind of uncertainty. The detailed codes within the more generic codes were compared and grouped into sub-codes if they were similar on a particular aspect, and different from other codes. As a final step all the stories were read again, and assessed on their fit. Stories were recoded when necessary. In the coding process, tensions in the emerging patterns were explicitly searched for, which led to re-examinations of the stories and interpretations. Table 2.1 displays the final code tree.

The sub-themes are distinguished based on the conditions under which they occur (see appendix 1 for coding table). Within the remainder of this article the main patterns and most exemplary stories will be presented and discussed.

4. The researcher's unfamiliarity and lack of substantive knowledge about auditing, could have affected the way respondents told their stories. Many respondents asked whether their stories were too detailed or difficult to understand, and whether they needed to give more explanation. Sometimes this was done, and at other times this was not necessary in order to understand the underlying message. It could also have affected the level of substantive detail with which stories are told in the first place. However, there are no reasons to assume this has affected the kinds of uncertainty that have been put forward in the stories. Also, respondents were encouraged to tell the stories as they wished to, and that if clarification was needed, the researcher would ask for this.

Table 2.1. Coding table with number of stories

Information Uncertainty
Finding proof (4 stories)
Interpretation Uncertainty
Experiencing dilemmas (8 stories)
Determining right decision (4 stories)
Action Uncertainty
Negotiating with clients (3 stories)
Responding to client's emotions and private situations (10 stories)
Impact of client's private situation on interaction (4)
Impact of client's emotions on interaction (5)
Intrusion into official's private life (1)
Encountering deviations from normalcy (8 stories)
Dealing with not cooperating entrepreneurs (2)
Dealing with complex or messy bookkeeping records (3)
Dealing with the legal logic (3)

2.5. Findings

Finding proof

The first general theme refers to a lack of information. By creating an account or storyline supported by sufficient evidence, tax officials equip themselves for the negotiation with citizen-clients, and when it gets that far, for the judicial trial. Tax officials ask themselves whether their account of the situation remains standing against the arguments and evidence raised by citizen-clients and/or their accountants. The following story aptly shows that the process of finding proof is highly intertwined with constructing a convincing account.

‘How are we going to prove he’s withholding a part of his turnover?’ (Respondent 7)

‘Well, with this hospitality business you actually need several sources to say that someone’s bookkeeping is not true. (...) How are we going to prove that he’s withholding a part of his turnover? Yeah, then you are... you actually look at the process of [the moment] the customer comes in and places his order. What is done with the order by the servers? How does it go through the process of this hospitality business? What did the entrepreneur say about this? Well, where could it have gone wrong? (...) What if he says he registers everything he has sold? Well, then the cash register tells you what he’s sold. What can we do more? We have data about related businesses and their gross profit ourselves. How can we, then, further prove? Well, then we’re going to do a third-party investigation. Is that difficult? Maybe not, but you have to try to go through the process and look at what happens, at where we can find a starting point for the substantiation of our numbers, and for the position the entrepreneur adopted. Or for the thing we expect that could possibly be wrong. We try to quantify that. And yeah, to create a calculation that is clear, and understandable, for both him and us,

with which... well yeah, we can go to court in the end. Where the judge again also says like okay, tax administration, you clearly and credibly mapped this out, and made it plausible that the entrepreneur committed fraud. Yeah, how difficult is that... (....) Yeah [it's about finding] the strongest proofs, which apply to the matter of course, but which you can also create, and which are communicable, also to the entrepreneur. It has to be understandable to the entrepreneur. It doesn't make any sense to make something... a nice mathematical formula, for instance, of which you as a mathematician think it's evidence, but don't get it explained and understood. Well yeah, then it stops. You can better take something which the entrepreneur has an understanding of, and use that, and his own stories, to provide rebuttal proof, to support your statement that he didn't register everything of his turnover. To create another picture that is understood by the entrepreneur.'

This problem of information is a particular one. The story shows that the respondent did not search information out of a sense of uncertainty regarding what happened. In fact, the respondent strongly felt that his account of the situation was true, because he had cues to believe so. The uncertainty rather consists in not knowing whether one is able to find enough evidence to substantiate one's account of the situation. This story shows how the respondent looks for more proof and tries to make a persuasive account with the collected evidence. Looking for more proof, for example, by means of third-party investigations, is also a common response in other stories within this theme. The emphasis is on constructing a convincing report with sufficient proof. Some respondents argue that when they are in doubt about whether they have a strong enough case, they sometimes satisfice and give their report a try, to wait and see whether entrepreneurs will object or not.

Ambiguity of interpretation

Interpretation problems, the second theme, can be distinguished from information problems. Whereas more information helps respondents in substantiating their account, it does not necessarily help them when they face situations that are difficult to interpret. As the following stories show, interpretation problems do not arise because there is an absence of information, but because the standards officials (can) use to evaluate cases are conflicting or vague in themselves. Two sorts of interpretation problems appear to be at play in tax officials' work: dilemmas and grey area interpretations.

Dilemmas

Stories about dilemmas mostly come down to a tension between what one ought to do as a tax official and one's personal values or ideas about what is appropriate, or one's feelings of empathy. This happens in different ways: firstly, when officials have to make decisions that have a profound impact on citizen-clients' lives, which challenges their personal feelings of empathy. Secondly, when officials have to make decisions that let entrepreneurs go on as usual, whereas they actually feel they should be put to a halt. The following story is an example of the former. The respondent in this story discovered that the entrepreneur did not keep mileage records of his lease car, because his accountant told him he did not have to.

‘Then I really have to leave it behind me’ (respondent 5)

‘The entrepreneur called his accountant, and they had a tough conversation. The accountant is of course always immune, because he does it [the records] for the entrepreneur. The entrepreneur is always responsible for his own bookkeeping records and tax return. Yeah, the entrepreneur also felt it was going to cost him a lot of money. So it wasn’t a nice conversation between the entrepreneur and the accountant. (...) The annoying part is that it was on the Friday before my vacation, and it already had taken a long time before I could make an appointment with him. I’m still in the human side of it, occasionally, because I of course brought him really bad news. Then I really have to leave it behind me. I knew he was also going on vacation that Saturday. The only thing that can work positively for him, is that he had a black box installed in his car. But eh, he never made a printout of it or whatsoever. So I said to him that he could maybe contact that company to request that data, and maybe puzzle with it. That’s the only thing we could still do. But that man wasn’t amused with his accountant. The entrepreneur called me that day to say he talked with the accountant and they’re going to try to get that data. But he said he was going on vacation with a strange feeling. Then I notice I really have to leave it behind me, that I have to think: “okay, it went wrong, and he’s going on vacation, me too, and I have to let go of it. It’s his problem.” I mean, (...) yeah, you just have to do your job. If you see something is going wrong, then you shouldn’t be pathetic about it. That’s the thing I run into occasionally, like, come on, this is the way it is, and not like that lovely motherly, worrisome feeling of how are we going to fix this for you.’

The story shows how the respondent struggles between following her feelings of empathy and carrying out her job. The ‘human side’ the respondent talks about makes her see different aspects of the case, than when she looks at it from the perspective of the law enforcer. The different interpretations do not compete over becoming the most truthful account, since both are obviously true for this respondent. In fact, they are based on different notions of fairness; one sticking to the law and highlighting the ‘objective’ error, the other shedding its light more broadly, also highlighting the entrepreneur’s private situation and mitigating circumstances. The story furthermore shows how the respondent follows the formal route and tries to look at it ‘professionally’, but still finds a way to cater to her feelings of empathy.

The other stories within this theme, likewise, show how respondents’ feelings of empathy lead them to also look at entrepreneurs’ private situations. However, most respondents argued they decided – sometimes after consulting colleagues – to stick to the formal route, and to look for other ways to help entrepreneurs.

Grey area interpretations

The grey area of rules and legislation elicits a second kind of interpretation uncertainty, which entails the question what is the right interpretation of the situation at hand. In the following story, the tax official not only problematizes this room for interpretation, but also the consequences in terms of differential treatment of similar cases.

***‘That’s easily said by the minister, good is good enough.
But what is good enough?’ (Respondent 13)***

‘We have the policy of good is good enough. That’s also the difficult area of tension of frontline inspectors. That’s easily said by the minister, good is good enough. But what is good enough? Letting go ten thousand euros? Is it [laughs] a thousand euros? That’s what makes it difficult for us. He could better have said: “let it go to a five thousand euros difference.” (...) Because you can better say, very strictly, just as we do with kilometers, that 501 kilometers is too much for the private use of a car. That’s very clear to us. That’s also very easy to inspect. (...). While lately, I had an employment services case... I think we really favored him (...). Yeah, that’s a difficult matter; they were classified in the wrong sector, which saved them almost sixty percent of premium to the tax administration. And they had a wage bill of a twenty million. Do the math; hundred-and-twenty thousand euros a year. I could have corrected four years. Then they would have been bankrupt. Five hundred people are working there... yeah, is that what you want?’

This respondent, thus, experienced difficulty in determining what is good enough, which is part of the new policy that encourages officials to look at the acceptability of tax returns. He continued his story by telling about a case which was favored. Later in the interview he said that he struggles with the question: ‘why do I correct in one case, and not in another? You always have that struggle’. The ‘grey area’ in rules and regulation thus engenders difficulties in determining the right decision, but – afterwards – also leads to personal struggles about whether the decision is fair in terms of equality. Within the grey area, this respondent argued, ‘you have to make sure you have someone with you, or two, three persons, with whom you deliberate’. In the absence of clear standards about what is right in these instances, stories show how colleagues are consulted to come to a decision that is supported by colleagues.

Action uncertainty

Officials’ stories show how difficult social interactions often boil down to (a fear of) losing control over situations or not knowing how to proceed to get control over situations. Within respondents’ stories different factors are associated with such uncertainties: unexpectedness of the situation, necessity of immediate action, and deviance from ‘normal’ situations. In situations with one or more of these factors present, officials often need to improvise on spot and/or need to change their plan of action. Officials are, thus, uncertain about their actions: they feel put on the spot, cannot carefully think through and prepare their actions, or their prepared plan of action does not suffice. Officials feel they (might) lose control over the situation and the outcome. Officials’ stories about action uncertainties are clustered around three sub-themes: emotional labor, negotiations, and deviations from normality. In what follows, the sub-themes will be discussed by using exemplary stories.

Emotional labor

A problem often referred to in tax officials’ stories stems from the fluid boundary between citizen-clients’ work life and private life. Tax officials share the basic expectation that the interaction with citizen-clients will occur ‘professionally’, that is, that it will center on the citizen-client

as entrepreneur who, in turn, approaches the tax official as a professional in performing the inspection task. When an interaction unfolds professionally, tax officials know what to expect and do. However, when 'private life' leaks into the encounter, tax officials experience this as difficult because they need to make immediate assessments of the situation and manage emotions. Respondents' stories show that this boundary is crossed when entrepreneurs' private situations or emotions influence the course of an interaction, and when entrepreneurs get personal at tax officials. The following story shows how an entrepreneur's unexpected emotional reaction leads to a deeply discomforting experience requiring on-the-spot emotional labor.

'How am I going to deal with this?' (Respondent 16)

'The strange thing was, I called that man, and normally you have a conversation, well, but this person was ranting on the telephone. Really, like, "what are you up to, why do you want to come, and I don't want you to come", you know. Well for me it was, even with my experience, yeah... I was sitting in a room with nine other colleagues, it was really busy... I hadn't anticipated on this. I thought, like, "how am I going to deal with this?" So I let him blow off steam, and tried to get the conversation uh... under control again. That has to be feasible with my experience. I walked out of that room to another room to be able to have a calm conversation, and to assess what actually happened. Because I didn't expect this at all, I never experienced that. (...) In such a conversation you really have to switch a lot; what am I dealing with, why does he react like that? Nowadays we work a lot with email. So I tried to make that step and said "if you give me your email address then I send you a confirmation, then you also have all my details (...) so you can check that if you want to." (...) Well yeah, at some point I had the conversation under control again. I managed to get in conversation with him again and to get his email address. That's an important step nowadays, because then you can reach him and he cannot flee anymore. (...) Well yeah, I was a bit perplexed by it.'

This story shows the unexpectedness of the entrepreneur's emotional reaction, the loss of control and the deeply felt discomfort. The respondent stresses the importance of getting the conversation under control again, and the action uncertainty accompanying that. The other stories show that such action uncertainties are mostly experienced on-site, where some respondents stayed at the inspection site and tried to make the situation 'workable' again, and others immediately left because they felt (sometimes physically) unsafe. In making such improvisational assessments and judgments, officials sometimes later on discover, and struggle with the fact, that they made a wrong decision.

Negotiating

Officials sometimes need and are even encouraged to negotiate with entrepreneurs, when there is disagreement about an interpretation of the law, or about how much money has been withheld, and the official does not have enough proof for his/her suspicions. The stories about negotiations bear witness to an uncertainty regarding assessing the other party's stance and willingness to accept the tax official's interpretation, and the discomfort it brings when the negotiation does not happen as wished. Next story shows the struggle the official experienced because he lost control over the situation.

'In retrospect I think we shouldn't have been so quick in... it's true, you always learn' (respondent 13)

'Sometimes I think like yeah, maybe we've been too accommodating in that world [of employment agencies]. Because we sometimes have like, okay, let's do this, we already correct so much. And afterwards in the final conversation it doesn't go like you... and then you think in retrospect: "damn, if we only had corrected that as well, because he doesn't actually deserve it." But you of course already made your decision. And you certainly keep struggling with that. I cannot deny that, you always keep struggling with that. (...). Yeah, I'm certainly not the only one, I'm convinced of that. (...). That's what we experienced with that big entrepreneur (...). We've given away quite a lot; we decided that, and expressed it, and now, afterwards, they are getting difficult about the corporate income tax. Maybe it's justice... In retrospect, we maybe shouldn't have been so quick in... that's true, you always learn. Maybe it would've been more convenient to have waited with real statements until the final conversation, then we would maybe have had some change. Like, "okay, we give you the sector premium, we all accept this, but the corporate tax to the contrary..." Then I would have had a better feeling maybe. But now I actually think it has been really much from my side, in retrospect. While we actually have been so open (...) And that's the area of tension; I was maybe too quick in... Then I am struggling with it for a week. I think "damn", and am bothered by the fact I actually gave away something, and he's making a fuss about this other thing. Maybe I shouldn't have done it. That's the learning process. I wouldn't do this again, certainly if it's about much money. So next time, I'm not going to struggle with this.'

In negotiating, the respondent experienced uncertainty regarding the entrepreneurs' stance and willingness to accept tax official's claims. Even more prominent in this story is the discomfort the official experiences after having found out he has been too open and has given away too much already early in the negotiation. The respondent feels that his openness towards the entrepreneur is exploited for personal gain, which is experienced as utterly unfair.

Deviations from normalcy

Some stories were about situations that could not be understood or handled by relying on the 'normal ways' of understanding and working. Some cases are considered rather complex and chaotic, which make it hard for tax officials to get an understanding of what is exactly happening. The stories about such complex cases show the difficulty tax officials experience in disen-

tangling the operational processes of these businesses before they reach the core of their task, that is, assessing whether the accounting records are acceptable. One respondent describes this difficulty as follows: 'It's a difficult process, (...) it's not a little project of which I know exactly what to do and where to look at tomorrow. With this kind of projects, it's complex' (respondent 11). Although officials often have more time to work on such cases, stories show officials do not know exactly what to do to get an understanding of such complex situations, because they already figured out (sometimes after having done quite some work) that their 'normal' inspection approach fell short. This sometimes involves a realization that officials have been doing work that should have been done by the entrepreneur, and that complexity might have been deliberately created to make things difficult. Stories show how entrepreneurs are dependent on the input of entrepreneurs and their advisors to get an understanding and to do the actual inspection: 'You have to make sure the advisor is collaborating. And that's the difficult thing, because he gets paid very well by the entrepreneur' (respondent 11).

The other stories within this theme show how entrepreneurs witness deviations from 'normal' behavior, and need to change their approach in order to deal with this. These stories are about entrepreneurs who do not cooperate, and about foot-dragging entrepreneurs who solely look at the legal aspects of an inspection rather than at the actual content. In such interactions, officials are 'on guard' and closely monitor their own actions, because of the fear to get sued:

'Then you have to pay attention to how you're going to maneuver' (respondent 8)

'They've hired two expensive guys from Deloitte, yeah, then you have to pay attention to how you're going to maneuver. (...) I don't have much problems with the fiscal side. But they are focusing on the formal side; are you acting correctly, and is the principle of an adversarial process respected? They are really much on the formal side of a correction. So if it's not going well, and it finds its way to court, then they'll say: "it's all well and good, but according to the law, this should have been done". Yeah, then your correction is gone.'

These advisors thus focus on how an inspection is done, which conflicts with the respondent's common focus on the content of the inspection. Another respondent also argues that such people are 'only looking at the legal aspects, and (...) not looking at what we are correcting' (respondent 13). Officials feel their on-the-spot actions are under close scrutiny, and can have major consequences for the course and outcome of an inspection. Table 2.2 provides descriptions of the different uncertainties found.

Table 2.2. Description of the kinds of uncertainty at play in frontline tax officials’ work

	Information uncertainty	Interpretation uncertainty	Action uncertainty
Problem of ...	Proof	Standards	Control
Contexts in which they occur	Lack of evidence to support one’s interpretation	Vague rules and legislation	Impact of citizen-clients’ private lives and emotions
		Conflicting norms, values, feelings	Negotiations with citizen-clients
			Deviations from normality
Difficulties experienced	Vague stories of citizen-clients	Law insufficient as backing	On-the-spot reaction
	Conflicting informational cues	Potential inconsistent decision making	Consequentiality of official’s immediate reaction
	Comprehensibility of account is not clear-cut affair	Far-reaching consequences for citizen-clients	Change of inspection approach
	Finding proof requires effort and time		Dependence on citizen-client

Exploring the sources of uncertainty

As a last step, the sources of the different kinds of uncertainty were explored to gain insight in possible explanations. Respondents’ stories show that different factors underpin and sometimes reinforce uncertainties. Some factors are inherent to frontline work in general, some seem specific to tax officials’ work, some to the policy of ‘horizontal supervision’ and others to societal changes.

A question about the truthfulness of entrepreneurs’ accounts of reality is underlying officials’ focus on substantiation and proof. The assumption is that there is an underlying truth, which can never be fully known. One respondent argues: ‘you never know how much black money is involved, it’s always a guess. Yeah and we have to prove it’ (respondent 17). Another respondent explains what makes it hard: whereas you generally inspect what is there, black money is about inspecting ‘what isn’t there, but which should have been there’ (respondent 16). This respondent also tells one can never be sure about whether an entrepreneur is ‘playing’ or not. This question is probably inherent to frontline work, and more pressing in situations where officials, such as tax officials, need to rely on citizen-clients as a source of information. This unfathomable character of ‘the truth’ could explain why respondents are not necessarily pre-occupied with finding out every detail of ‘the truth’, but with finding proof and constructing convincing accounts.

Room for interpretation of rules and regulation is inherent to discretionary frontline work, and sometimes involves ambiguity when these rules need to be interpreted in specific situations. In assessing specific cases against standards, officials sometimes question such standards when large deviations are encountered: ‘Could that standard be too generic, and could it [the

case] deviate?’ (Respondent 16). Also when it comes down to determining the height of fines, officials emphasize they have much leeway, which they sometimes struggle with: ‘It’s actually all wrong, but the correction is already very high. Should you then also impose such a high fine? It may be stated in the law, but we have the freedom to struggle with it’ (respondent 13). Stories have shown that this leeway or ‘freedom to struggle’ involves dilemmas between following the law on the one hand and feelings of empathy on the other hand. Whereas interpretation leeway is inherent to discretion, within the policy of ‘horizontal supervision’ it is enlarged and seen as part of professional judgment. Tax returns are not simply good or wrong, but acceptable or not, and the standards to assess this acceptability are purposefully left more open and up to the frontline official to assess. These stories show that this is not always an easy task and such grey areas involve struggles about what is right. This horizontal policy, furthermore, encourages officials to make settlement agreements with entrepreneurs when there is disagreement about an interpretation of the law, or the amount of money that has been withheld. Officials’ stories show how negotiations emanating from this are experienced as deeply discomforting when officials feel they (may) lose control over the situation and the outcome, or when they feel the resulting outcomes are not fair. The horizontal policy, thus, puts more weight to officials’ on-the-spot actions and negotiation skills, involving action uncertainties.

The indeterminacy of interactions with clients most prominently emerged as a factor underlying tax officials’ action uncertainties. How an interaction unfolds is only predictable to some extent; based on repeated experiences tax officials form certain expectations of how interactions normally unfold. However, tax officials acknowledge that social interactions with citizen-clients are indeterminate: ‘you’re on completely unknown territory, and ... you have to wait and see what’ll happen inside’ (respondent 3). This social uncertainty is inherent to frontline work, and interactions are sites where officials possibly lose control. Action problems within complex cases, moreover, often involve information and interpretation uncertainties. In complex cases, respondents commonly lose ‘overview because of all the details’ (Respondent 6), and do not know what to do exactly in order to get this overview, and need to rely on entrepreneurs’ input. The stories about such complex cases show that uncertainty about how to get a grip on such situations is reinforced by a lack of insight in the operational processes of a business, ‘shortcomings’ of the law, and absence of the jurisprudence.

Lastly, respondents acknowledge that aversion towards the tax administration on the part of entrepreneurs is inherent to their work, because of negative prior interactions or a belief that tax officials only come to get money. Therefore, tax officials are mindful of entrepreneurs’ emotions. Some respondents associate emotional interactions and foot-dragging entrepreneurs with hard economic times. One respondent argues, for instance, that entrepreneurs in a particular sector are foot-dragging and committing fraud because ‘they cannot do otherwise, because their prices are under pressure’ (respondent 13). Another respondent holds that he already knows the situation is not going to be nice when someone has ‘bad figures, and the economy is getting worse’ (respondent 3).

2.6. Conclusion and discussion

In line with the rational view on bureaucracy, tax officials experienced problems of information. They, however, were not in the dark about what happened. Rather, they already seemed to know very well what was the matter, and looked for proof supporting their account. In accordance with uncertainty as discussed within street-level bureaucracy literature, tax officials also experienced interpretation problems. Interpretation problems entail difficulties with standards

Chapter 2

to make decisions, i.e. conflicting feelings and norms and vague rules and regulation. These yield questions as to what is right. Tax officials indicate they deliberate with colleagues in case of doubt, in order to have a back-up and to struggle less with it afterwards. This study also found that tax officials face action uncertainties. Action problems involve uncertainty about control, i.e. (a fear of) losing it in interaction with the entrepreneur, or about how to get it (back). This often happens in interaction with entrepreneurs and their businesses, which can never be fully predicted. This puts officials on guard because they often need to take ad hoc action and feel their on-the-spot actions weigh heavily on the course and outcome of an interaction.

This study also explored the sources of these uncertainties and has shown that there are structural aspects of frontline work, such as the 'unknowable truth' and ambiguous character of rules and regulation, which underpin officials' search for proof, and uncertainty about what is right. In addition, stories also point to the 'horizontal supervision policy' as a source of different uncertainties. It gives frontline officials more room for interpretation, which sometimes involves doubt about what is the right decision. This policy furthermore encourages officials to negotiate with citizen-clients, where officials feel they possibly lose control. Economic hardships and the undesirability of audits on the part entrepreneurs underlie the latter's emotional outburst in interactions.

These findings underline the importance of social interactions to bureaucratic work and hence to understanding the role of uncertainty in bureaucracy. Whereas public administration literature has pointed to the existence of information uncertainties (e.g. Simon, 1976), and interpretation uncertainties (Lipsky, 1980), this study adds a third kind: action uncertainties. These uncertainties are about how to maintain or get control over situations (see Dubois, 2014, on control practices), and can generally not be resolved by gathering more information, relying on rules and consulting colleagues, but require officials to make immediate assessments and judgments. This raises interesting questions with regards to the scholarly writing on street-level discretion and control. Much research on street-level bureaucracies has been done within what is depicted by Maynard-Moody and Musheno (2012) as the 'implementation-control-discretion narrative', focusing on how street-level bureaucrats make decisions within given rules, and rely on abstract knowledge. This current study shows that some frontline situations are, by their very nature, difficult to control by street-level bureaucrats, let alone by upper-level managers. An important line of future research, then, is exploring how frontline officials improvise and rely on practical knowledge in face of uncertain situations (see Maynard-Moody & Musheno, 2012; Wagenaar, 2004). What are the beliefs, understandings, values and even emotions officials rely upon to deal with these action uncertainties?

Moreover, whereas the street-level bureaucracy literature has a tradition of explaining frontline discretion, pointing to the influence of characteristics of the work context, frontline officials and citizen-clients, this study points to the importance of a more dynamic understanding of frontline decision making. Research approaches should allow for the study of interaction processes, citizen-clients' perspectives, and their negotiation and communication skills (e.g. Bartels, 2014), especially in policy contexts where interpersonal notions such as trust are emphasized. Although such policies may yield more responsive law enforcement and service provision, they could thus also compromise consistent and fair decision making, especially when certain types of citizen-clients have better negotiation and communication skills to take control in bureaucratic interactions.

The findings also have implications for the accountability and public management literature. This study has shown that a policy that gives more leeway to street-level officials could constitute a source of uncertainties at the frontline. The act of categorization is delegated to

individual officials who, to recall one of our respondents, have the ‘freedom to struggle’. It practically means that officials have to work with rules and legislation that are vaguer, and that they are encouraged to negotiate with citizen-clients when there is disagreement in ‘grey areas’. This study has shown that officials often deliberate – on their own initiative – with colleagues to deal with interpretation uncertainty and to have ‘a backing’, which is in the literature also referred to as ‘professional accountability’ (Hupe & Hill, 2007). Future research should focus on whether and how frontline officials working with such horizontal policies, collectively deal with uncertainties, and whether and how management has an influence on these deliberation practices (see Piore, 2015; Rutz et al., 2015).

A limitation is that this study solely used the storytelling method. Whereas the rich narratives yield insight in the uncertainties experienced and the meanings attached to these, they are less apt to study officials’ actual behavior. This study has given insight in how officials meaningfully deal with uncertainties experienced, and offered a first step in showing possible explanations. Future studies could complement this effort by conducting participant observations and experiments to study officials’ behaviors under different conditions. Another limitation of this study is its focus on a particular type of street-level bureaucrats. This study has yielded insight in the uncertainties of street-level bureaucrats who have much discretion, a lot of information about citizen-clients, and who have come to work with less strict rules and regulation. It is likely that street-level bureaucrats with less discretion, less information available and ‘a never-ending demand for more and better services’ (Nielsen, 2015, p. 117), experience these uncertainties differently and with a different importance. Comparative research is needed to develop a theory on uncertainty at the frontlines. To this end, future research could compare uncertainties experienced within different policy fields and different types of street-level bureaucrats to theoretically advance this field of study.

Chapter 3

From poker games to kitchen tables How social dynamics impact street-level decision making

Abstract

Existing research on bureaucratic encounters typically studies how bureaucrats' and clients' characteristics influence frontline decision making. How social dynamics in interactions between street-level bureaucrats, and between officials and citizens may impact case-related decision making, largely remains an underexplored field of study, despite the fact that new forms of governance introduce social dynamics in the form of trust, collaboration and responsiveness as tools to increase legitimacy. Comparing in-depth qualitative data of the Belgian labor inspectorate and the Dutch tax authorities, this study scrutinizes how frontline social dynamics may affect officials' decision making about specific cases.

Raaphorst, N., & Loyens, K. *(Under review)*. *From poker games to kitchen tables: How social dynamics impact street-level decision making*. Submitted to an international peer-reviewed journal.

3.1. Introduction

The relationship between frontline officers and citizen-clients has always been at the core of street-level bureaucracy literature (Bartels, 2013). It has gained importance with the introduction of new public management reforms – aiming to increase performance and legitimacy – that have led to the reformulation of the bureaucratic encounter as a businesslike arena, where officials are responsive to citizens' (or clients') needs and demands (Vigoda, 2002). Citizens then become more involved in the process of public service delivery. Post-NPM movements, based on notions of participatory governance, even point to a more horizontal relationship, where officials collaborate with citizens and other stakeholders to deal with complex problems (Bartels, 2013; Vigoda, 2002). The process and outcome of frontline decision making has hence come to depend on interactions between officials and citizens (Bartels, 2014). Such trends are not only visible in public service delivery, but also in public organizations preoccupied with law enforcement like tax and inspection authorities (e.g. Mascini & Van Wijk, 2009). Officials are for instance encouraged by their organizations to negotiate with citizen-clients. This brings the social interaction between officials and citizens to the center of frontline decision making. There is, however, little insight in how case-related decisions are made as part of interactions with citizen-clients, i.e. how the social dynamics of such an interaction affect decision making (Bruhn & Ekström, 2017).

Furthermore, street-level bureaucrats are embedded in professional networks with coworkers and peers as actors giving support and advice when needed (Brehm & Gates, 1997; Lipsky, 1980; Maynard-Moody & Musheno, 2000; Vinzant & Crothers, 1998). Nowadays, it is assumed that new service pressures such as multiproblem clients even require multiprofessional action (Noordegraaf, 2011, 2016). Therefore new organizational arrangements are arising that enable collaboration between different agencies and professionals, thereby avoiding fragmentation (Groeneveld & Van de Walle, 2011; Noordegraaf, 2011). In various sectors, such as health care and social services, programs are set up in which frontline workers across agencies work together to improve quality and availability of social services (e.g. McCallin, 2001; Sandfort, 1999). Inspectorates too engage in partnerships and work in multidisciplinary teams of inspectors on common issues, based on shared discretion (Rutz et al., 2015). Noordegraaf (2011) points to the complexity of such multiprofessional interactions, resulting from possible clashing professional cultures, difficulty in exchanging knowledge and other practical problems. Social dynamics between officials within such interdisciplinary teams are said to add to this complexity. Their impact on frontline decision making is, however, an underexplored theme within the literature, as is the role of 'collective discretion' (Rutz et al., 2015).

Thus in the call for more responsiveness, horizontal relationships with citizens and other professionals are introduced into public organizations, implying negotiation with clients and collaboration with peers. Whereas the street-level bureaucracy literature has described how frontline decision making is embedded in official-client and collegial interactions, it mainly focused on either client or official characteristics (e.g. Maynard-Moody & Musheno, 2003). What has thus far largely remained out of sight is the impact of social dynamics in which street-level bureaucrats operate, involving interactions with clients and peers, both in uni- and multidisciplinary teams (but see: Rutz et al., 2015). This is striking, given that such complex social dynamics in frontline decision making are becoming increasingly important in governance nowadays (Bartels, 2013; Yang, 2005). To address these gaps, this paper aims to answer the following research question: how do social dynamics within different frontline contexts impact case-related decision making? We specifically look at how aspects of social interactions or social relations are involved in decision making about specific cases. Comparing different social

frontline constellations of two inspectorates (labor inspectorate and tax authorities), and relying on the methods of storytelling and observations, this paper analyzes how social dynamics in citizen-official and collegial interactions affect decision making. These two cases are compared because they entail and allow us to compare different frontline social constellations.

This paper will bring the street-level bureaucracy literature further by (1) addressing the call for research on the role of social context and social dynamics in street-level work (Bruhn & Ekström, 2017; Siciliano, 2015), and (2) answering the call for more comparative research in street-level bureaucracy which is important to increase knowledge about frontline behavior in various contexts, but which is still scarce (Pollitt, 2013; Hupe et al., 2015).

3.2. The social context of frontline decision making

Within public administration, different streams of literature focus on the social context of street-level bureaucracy. We shortly discuss the socialization, professionalism and accountability literature, and argue that more insight is needed in how street-level bureaucrats' social context affects case-related decisions. Therefore a broader notion of social context is needed, also incorporating situational social dynamics.

The socialization literature shows that attitudes and identities of newcomers to public organizations become more homogenous over time (Moyson et al., 2017), by which they become organizational members (Oberfield, 2014). Officials learn the ropes of the job by relying on their (in)formal networks of colleagues for information and social support (Hatmaker et al., 2011; Van Kleef, 2016), and communication with colleagues increases uniformity in law enforcement (Van Kleef, 2016). In line with this approach, recent studies have begun to study the role of social network structures and composition in street-level innovation and performance (Maroulis, 2015; Siciliano, 2015). The professionalism literature adds that the transfer of knowledge and expertise is facilitated by professional associations, institutional arrangements and shared education (Schott et al., 2016). Studies have also concluded that professional norms shared by frontline officers affect street-level decision making (Sandfort, 2000; Maynard-Moody & Musheno, 2003). Both literatures offer valuable insights in, respectively, the impact of social networks on learning processes of newcomers and the role of professional norms. However, their scope is too limited for the purpose of this study, given their focus on the homogenizing influence of peers or on similarities across officials. The focus is mostly on the structural aspects of peer networks and on outcomes such as performance, innovation and learning. The dynamic aspect of social interactions and actual decision making are not taken into consideration.

The accountability literature broadens the perspective by not only emphasizing the role of vocational associations, peers and colleagues, but also other actors surrounding street-level bureaucrats. Hupe and Hill (2007) mention three fora of public accountability, being (1) professional accountability (also addressed in socialization and professionalism literature), (2) public-administrative accountability, including performance indicators, managerial oversight and the rule of law, and (3) participatory accountability, including citizens and interest groups. These fora form a 'complex institutional web' (Hupe & Hill, 2007, p. 290) in which street-level bureaucrats are confronted with different, possibly contradicting, demands from actors they are accountable to (Behn, 2001). In specific situations, street-level bureaucrats either try to find a workable balance between them (Thacher & Rein, 2004), or consider the demands as possible repertoires they can choose from, thereby strengthening their discretion. This literature thus provides a general framework of actors to whom street-level bureaucrats are accountable, but does not offer any insight in how frontline decision making is shaped by these accountability

fora, and to whom street-level bureaucrats feel accountable (if they indeed do so). Hence, more insight is needed into the underlying processes that explain how actors surrounding street-level bureaucrats shape frontline decision making.

Some studies do offer some clues as to how social dynamics between officials and citizens and between coworkers may impact on street-level decision making. The study of Bruhn and Ekström (2017) on frontline interactions, for instance, shows how decisions are designed in interactions with citizens and how ‘the progression of the conversation is conditioned by the client’s acceptance of the decision’ (p. 206). Furthermore, they show that even in a street-level context with strict rules, these rules are negotiated in concrete interactions with citizen-clients, and argue that the application of rules should be seen as ‘interactional achievements’ (p. 199). Pointing to the more structural aspects of interactions, Nielsen (2007) shows that street-level inspectors’ treatment of citizens being regulated is dependent on characteristics of the interaction; the higher the level of negotiation and cooperation, the more lenient the inspector’s enforcement is. That study points to the relevance of the frequency and quality of interaction for how citizen-clients are treated.⁵

With regards to interactions among coworkers, Keiser (2010) shows that, even without a high level of interaction between street-level bureaucrats, individual street-level bureaucrats’ decision making is shaped by their perceptions of how other colleagues would make the decision. It is held that expectations or knowledge that other organizational units would overturn particular decisions, discourages street-level officials to make such decisions. Furthermore, in line with street-level bureaucracy studies, Rutz et al. (2015) show that inspectors working individually sometimes consult colleagues for advice or specific know-how. They also show that teams with shared discretion have a greater action repertoire and mandate, because inspectors can make use of other inspectors’ authority and abilities. That study mainly shows how inspectors purposefully rely on coworkers to act both consistently and responsively.

This current study builds on these studies by exploring the impact of social context, including more situational social dynamics, on decision making about specific cases in different frontline social constellations.

3.3. Description of research contexts

This paper shows the results of two case studies, respectively conducted in a Belgian labor inspectorate and the Dutch tax authorities. The central aim is to increase knowledge of street-level bureaucracy in various social contexts. The focus is on cross-context, and not cross-country comparison, exploring how interactions with citizen and peers affect frontline decision making.

5. In line with street-level bureaucracy scholars arguing that interactions between street-level officials and citizens are underexplored and deserve more scholarly attention given the centrality of these interactions in the public sector (e.g. Bartels, 2013; Bruhn & Ekström, 2017), Nielsen (2007) contends that, in the regulation literature, the interaction between inspectors and inspected parties is mostly treated as a black box, mixing characteristics of inspectors’ actions with characteristics of interactions. How interactions are performed is usually conceptualized and measured as part of inspectors’ enforcement styles, which can be strict, lenient or both. Studies on explaining enforcement styles generally focus on external factors such as the organizational setting, political environment and the attitudes of the inspector (Nielsen, 2007; May & Winter, 2011).

Case study in a Belgian labor inspectorate

The first study is conducted within the Belgian Inspectorate of Social Laws, which inspects undeclared work violations and enforces wage and labor condition regulations. In undeclared work inspections, inspectors consult the social security registration system to check which employees have been registered by entrepreneurs and then make 'on-site' visits to identify employees who are present. Wage and labor condition inspections mostly start with a complaint (for example, from a (former) employee who did not receive the correct wage), after which inspectors go 'on-site' to examine pay scales, reimbursements and labor regimes. Inspectors who detect violations have the discretion to make a report to law enforcement, to propose a settlement agreement or to give a warning. Managers in regional offices supervise decision making to maximize consistency within their office, but their authority is not always recognized. Encouraging consistent decision making between regional offices, is not considered a priority.

Inspectors in this agency work alone, with a colleague (in complex or sensitive cases) or in interorganizational teams. They do inspections within a specific district, allowing them to build a strong working relationship with the entrepreneurs there. Inspectors are involved in two types of interorganizational team work: (1) anti-fraud teams with labor inspectors from the five Belgian labor inspectorates who focus on undeclared work violations only, and (2) interdisciplinary teams with representatives of different enforcement agencies, such as the police, the tax inspectorate, the food inspectorate and the labor inspectorate, in which each inspector focuses on his own specialty. While the former type of team cooperation is aimed at increasing consistency of decision making across organizations, the latter type aims at enhancing efficiency and decreasing the frequency of inspections. Despite interorganizational cooperation, each case file is dealt with by inspectors with individual discretion.

Case study in the Dutch tax authorities

The second study is conducted within the Dutch tax authorities, and focused on frontline tax officials who audit the tax returns and bookkeeping records of small and medium sized businesses. They have face-to-face interactions with entrepreneurs, mostly 'on-site'. To enhance compliance, the Dutch tax authorities have adopted a policy promoting responsive law enforcement (Gribnau, 2007) in which officials are encouraged to approve bookkeeping records instead of disapproving them (that is, by not correcting every gap, but making future agreements with entrepreneurs) and have more leeway to take into account contextual elements, such as the demeanor of entrepreneurs.

Depending on the complexity of the case, inspections take a week to months, and inspectors work alone or with colleagues. Inspectors generally have not previously seen or inspected the entrepreneur, but have a lot of information available about the case to be inspected. They have much leeway in interpreting cases, choosing their actions and making decisions. Some tax officials are part of the multidisciplinary 'take-away team' aimed at combating organized crime and confiscating criminal assets. It involves a collaboration between five large municipalities in the province of Brabant in the Netherlands, the public prosecutor, the police, Fiscal Information and Investigation (FIOD), the tax authorities, and the Royal Military Police. The team started in 2012 at two locations as part of a two-year pilot, which has been extended another three years, and to another province. The different parties collaborate to take away criminal assets as quickly as possible, thereby increasing legitimacy by showing that crime is tackled. The parties exchange information about suspects, and collaboratively make a decision.

3.4. Methodology

In both case studies qualitative methods were used. The labor inspection study had an ethnographic design using observations (of inspections and (in)formal meetings), informal conversations and semi-structured interviews with labor inspectors and managers over a period of 18 months. This paper draws upon a doctoral study on decision making of labor inspectors and police officers (Loyens, 2012). Four Flemish regional offices were selected that varied on size and type of districts (urbanized or not). Of the 18 inspectors and four managers who were interviewed, 11 were women and 11 were men. The average age was 45, varying between 26 and 60. The central goal was to explore how inspectors deal with and reason about dilemmas in decision making. Observation and conversation data were written down in detailed field notes, while formal interviews were audio-recorded and transcribed verbatim. For the purpose of this paper, the dilemmas in which client-inspector and collegial interactions were important for case settlement were selected.

In the tax authorities study, interviews were conducted with tax officials aimed at soliciting stories about situations they experienced as difficult, comparable to the dilemmas in the first study. The data used for this paper have been collected as part of the study on tax officials' uncertainty experiences (chapter 2). Stories are suitable for showing how street-level bureaucrats meaningfully deal with difficulties or tensions in their everyday work (e.g. Maynard-Moody & Musheno, 2003). The storytelling approach allows for gaining insight in how respondents give meaning to situations, showing what they find important. When studying on-the-spot frontline decision making, such approach is more suitable than probing for respondents' ideas about abstract concepts. The interviewed tax officials work in different tax offices across the Netherlands. Of the 17 interviewed respondents, 14 are male and three are female. Ten respondents have been in service for over 30 years, one for 18 years, and six have been in service for less than ten years. The interviews were audio-recorded and transcribed verbatim. The interview transcripts were reanalyzed, looking specifically at the role of social dynamics/contexts in decision making. This was done by specifically looking at how frontline officials experience or reflect on how aspects of social interactions are involved in decision making about specific cases.

To enhance intercoder reliability, the researchers both coded data they collected and a part of the data collected in the other case study. Interpretations of the data were discussed in four coding sessions, and subsequently refined. Even though the data for this study were collected independently and for different purposes, both studies are comparable because they contain inductive and in-depth data about how different social contexts influence frontline decision making. See table 3.1 for a summary of the case descriptions of both studies.

3.5. Findings

This part illustrates the different ways in which street-level decision making is shaped by social dynamics (summarized in table 3.2). Similarities across cases can contribute to an overall theory of social context in street-level bureaucracy.

Discussing and negotiating with the client

Within both organizations, settlement agreements are seen as a means to enhance entrepreneurs' compliance. Settlement agreements allow more leniency, provided that entrepreneurs agree to do it right in the future. In practice, it is often used in 'ambiguous' situations, where there is room for different interpretations or not enough evidence. Oftentimes officials need to negotiate with entrepreneurs to come to such an agreement.

Table 3.1. Description of the cases

	Belgian labor inspectors	Dutch tax officials
Inspectee	Small, medium and large sized entrepreneurs	Small and medium sized entrepreneurs
Discretion	In case of undeclared work or violation of wage and labor conditions law, they can make a report to law enforcement, propose a settlement agreement or give a warning	In case of fiscal gaps or violations of fiscal laws, they can impose tax corrections, fines and propose settlement agreements
Context of inspection	Interorganizational teams with narrow tasks (undeclared work)	Mostly individual officials, who sometimes involve colleagues
	Interorganizational teams with broad tasks (undeclared work and wage and labor conditions)	Interorganizational team with shared objective (combating organized crime)
Educational backgrounds inspectors	Vocational, higher professional and academic education in fields as economy, law, social work and criminology	Vocational, higher professional and academic education in fields as accountancy, economy and law

The following story excerpt underlines the importance of social dynamics for the final decision tax officials make:

‘Yeah, you have to prepare very well, and to discuss with specialists here at the office. You have to have a good story there. Because if you are there, and they have a rebuttal, and you start stuttering and stammering, then it’s not going well. Then it becomes difficult to still... because it’s a game, and they also feel if you’re not standing strong. Then they will open their bag of tricks. (...) So it went well. I prepared very well [laughs]. If something like that happens, I want to know everything of it. Then I’ll have a conversation with them, and if they come with things and we have no answer at that moment, yeah, then they did better. I’ve no problems with that.’

This story shows that tax officials’ decisions largely depend on the involved actors’ reactions and negotiation skills, in addition to the available evidence (which is used as leverage). The respondent even compares it to a game that has to be played out. This game, it is held, is about who has the most convincing account: ‘You have to convince the entrepreneur of your point of view. If they agree, then they agree. (..) Then it could be that they read a section of the law differently. That’s possible. Then they need to convince me.’ The labor inspectors also use the game metaphor (particularly poker) to describe interactions with entrepreneurs. In situations where they lack evidence to prove their case – which often occurs in presumed overtime violations – labor inspectors try not to show the entrepreneur that they are in the weakest position. To recover as much money as possible for the disadvantaged employee(s), a ‘poker face’ is essential, although fear and lack of legal knowledge on the part of the entrepreneur are helpful too. In both organizations, the underlying idea is that negotiation results in greater financial benefit than a sole reliance on available evidence would generate. How social dynamics shape this negotiation process is clearly shown in the following story from a tax official who received a snitch letter asserting

Chapter 3

that an entrepreneur fires his employees in the winter, letting them claim unemployment benefits, and, at the same time, letting them do undeclared work within the winter period.

‘But yeah, there we were, not knowing the amount; I missed a lot of hours, but I couldn’t prove it. You then try to convince them, based on facts and substantiations. So it resulted in a compromise, with which he [the entrepreneur] got away very well. But well, I was proven right. (...) It was actually his own fault. We got a snitch letter stating that those people received unemployment benefits, but proceeded their work as painters. So I was substantiating a lot, like: “You have one person on the payroll, and you paint yourself, and I have barely seen those self-employed people, so you did something with those people.” (...) I said: “Your income per hour in the winter is just ridiculous.” And then he said: “But the boys that help me out during the winter...” And I thought: “Now I got you!” I said: “You’re standing on thin ice, sir. (...) You’re facilitating benefit fraud. (...) If the UWV [employee insurance agency] knows about this, then you have a very big problem.” At that moment, he admitted it. So I said: “Tell me what you want: do you want me to correct the boys [employees], but then they have a big problem, because then UWV discovers it [laughs], and then they will also come to you, or we do it by means of the anonymous tax rate, but that is of course a much higher correction”.’

Although the tax official did not have enough evidence at the start of the interaction, the entrepreneur accidentally admitted the violation, because he was figuratively pushed to the wall. The decision that follows thus results from how both officials and entrepreneurs stand their ground and react to each other.

Labor inspectors also negotiate with employers, but not always in a way that is supported by the agency. The agency’s policy prescribes that labor inspectors make a report to law enforcement in the case of undeclared work, but encourage them to make settlement agreements in the case of wage and labor condition violations. However, in practice, labor inspectors regularly use undeclared work violations as leverage to make a settlement agreement concerning employees’ wage, as illustrated in the story below:

‘In [small shop], I also had a case with two undeclared employees who had not received any wage the past two months or so. So we made a kind of deal actually. The employer admitted that she had not registered the two employees, but she agreed to register them retrospectively for the past two months and pay them their wage, if we did not make a report to law enforcement for the undeclared work violation. So she corrected the situation. That is more important than being able to collect a fine [approximately 5000 euro], because she paid social security taxes for two months and these employees received their money.’

Bargaining with the employer is thus considered a strategy to find a solution that is in the best interest of the employee(s), as well as the treasury.

In sum, in negotiating with clients, final decisions are not only dependent on officials’ thoughtful considerations, but also on both parties’ on-the-spot persuasion and negotiation skills.

Informal social network to deal with subjectivism

Whereas tax officials view themselves as professionals who need a degree of freedom to adjust their enforcement styles to the situation at hand and to make informed decisions, they are sometimes concerned about inconsistencies in decision making, resulting from the subjective nature of their decisions. This subjectivism runs counter to their strong sense of justice, and is fostered by the idea that entrepreneurs in similar situations should be treated similarly. The following interview excerpt shows how a tax official experiences a tension between discretion and equal treatment:

‘We luckily still have freedom. But it remains something very subjective of course, because I’m inspecting with a digital dossier that I need to fill, but what I fill out, remains subjective of course. I can leave something out, or mention something. So if you have an enterprise and you say something to me I don’t think is very kind, then it can have an influence... yeah, it is not supposed to be like that, but it involuntary plays a role. It depends on your relationship, right? If I have a good relationship with you... well, then I’m the first to say that it is correct, and it influences how you treat someone. When there are points of difference, I think you’re much more inclined to approach each other. Then I won’t easily say like, I continue my inspection to find something. Then I’m more inclined to say well, okay, possibly there could be something wrong, but I close the book and we shake hands.’

Like other respondents, this inspector acknowledges that decisions are involuntarily influenced by relationships with entrepreneurs. This might result in more lenient or strict treatment. A labor inspector explains that an uncooperative entrepreneur in a recent case was ridiculing the inspector for not having enough proof; the inspector then pulled out all the stops (with house searches and confiscations), just to get back at him. A tax official explains that a bond of friendship can also influence one’s view of the case, because even though ‘we are government officials who are supposed to be as objective as possible’, ‘we are humans’. The inspector acknowledges that ‘[a]nother colleague would maybe have made different decisions’. These examples clearly show how inspectors experience that relations with citizens, i.e. feelings of friendship or hostility, could affect their inspection style and decision making, standing in the way of being ‘as objective as possible’. When this subjectivism is experienced as a problem, officials use different strategies to make decisions less subjective. Besides relying on overall inspection guidelines (tax officials), and official-specific ‘rules of thumb’ to safeguard equal treatment across cases (labor inspectors), officials also rely on their peers to reach a widely supported decision.

‘That is the hard part... when do you have enough assurance, when have you done enough work? Yeah, that’s hard to tell, that’s different for each inspection. In case of doubt, you need to consult with your colleagues, just briefly talk about it. (...) you get a judgment supported by colleagues. That’s enough then. Professional judgment, that’s what it is called.’

Consulting one’s colleagues is a strategy that is commonly used by tax officials. By attuning their decisions to colleagues’ opinions or assessments, officials feel their case is stronger and less subjective. One respondent refers to this practice as making ‘intersubjective’ judgments, which are shared among colleagues. Such consulting practices can also help to prevent ‘tunnel vision’

by letting ‘a colleague objectively assess whether you’re still on the right track.’ One respondent argues he looks for ‘likeminded colleagues’, who do not correct everything, and do ‘not want to go to extremes’. This suggests that ‘intersubjectivism’ is sought by consulting colleagues who share the same inspection attitudes and practices.

The labor inspectors are not so much concerned about reaching a widely supported decision. Only newcomers in the job regularly consult with other inspectors, but quickly learn the importance of developing a personal inspection style, as one of the newcomers explains:

‘There’s the style of always believing in the goodwill of the entrepreneur, and constantly excusing for doing one’s job and requiring so much of entrepreneurs’ time (...). Then there are those men who, rather straightforwardly, require entrepreneurs’ cooperation, or else they will make a report for obstruction of the investigation. (...) Those are the extremes, and other colleagues’ styles are in between. That shocked me in the beginning (...) I didn’t know how to behave myself, but then I realized that whatever style I would choose, it would be okay (laughs). If these extremes are accepted, then anything goes, right?’

Interorganizational teams with shared tasks: Same rule, different meaning and enforcement

The social context, and thereby also the decision making, changes when street-level bureaucrats are part of teams. The labor inspectors under study monthly participate in regional anti-fraud inspections, in teams consisting of inspectors of the five Belgian labor inspectorates. Given that inspectors are encouraged to always make a report to law enforcement in the case of social fraud, similar enforcement can be expected when labor inspectors work alone or in these uni-disciplinary teams. However, observations showed that differences in social context lead to different decisions. Particularly, social dynamics shape how the ‘strict enforcement’ rule is interpreted and enforced. A labor inspector explains that the law is strictly enforced in team inspections, but not in individual inspections, because then inspectors believe they can be more lenient.

‘The agreement in these anti-fraud teams is to adapt a kind of zero tolerance policy [consistent with agency’s policy]. This means that even if an employer registers the employee the day after the inspection, we make a report to law enforcement. In regular inspections (...), you can – and I have already done that – be more lenient by offering him to drop the charge if he registers the employee correctly afterwards. If he registers the employee, also for the day of the inspection of course, then I will not make a report. If he does not do so, I will make a report.’

Observations of anti-fraud teams and individual inspections confirmed that the social context surrounding the case shape the meaning and weight of the ‘strict enforcement’ rule. In team inspections, inspectors for example ask the question ‘Who will make the report?’ rather than ‘Who will take this case?’ when violations are detected, implying that making a report to law enforcement is considered the only acceptable response. In interviews, labor inspectors explained that, when working in an anti-fraud team, the ‘strict enforcement’ rule is seen as a formal instruction that should be followed or, as mentioned above, an agreement they have made with other agencies. However, if they work alone, individual discretion prevails and the same

rule is considered a mere guideline or recommendation. The different interpretation could be explained by inspectors trying not to appear 'soft' in front of colleagues from different agencies, as is illustrated by the following interview excerpt, where two inspectors from the same agency make a secret deal with an employer during an anti-fraud inspection, because they think he deserved leniency:

'If there are, for example, eleven employees of whom ten are registered and one is not, then I dare say to the employer: "Write his name down on the list as well." But I will not tell the others. I can tell my colleague [person working in the same agency], but not those guys from the [other labor inspectorate], because they will start bossing me around and tell me I have to make a report. But if two inspectors of our agency are present, we can make a deal. (...) We do that without letting the others know. We do it sneakily.'

Also the more distanced relationship between inspectors and clients in team inspections could explain the difference. In individual inspections, inspectors might deliberately use leniency to foster the long working relationships with entrepreneurs and thereby aim to increase compliance.

Multidisciplinary shared discretion: Informal agreements

Dutch tax officials also participate in interorganizational cooperation. The so-called 'take-away team' involves professionals from different agencies (including the tax authorities) to combat organized crime and confiscate criminal assets. To enhance cooperation, they have found a way to collaboratively work around laws that are experienced as restricting the possibilities to combat crime. In so-called 'kitchen table conversations' confidential information is openly shared:

'Every Tuesday we have the kitchen table, that's how we call that. All things that are brought to the attention are discussed (...) in an informal way. Because we have the obligation of confidentiality. If a police officer asks me: "We have that person in the picture, could you check what kind of earnings he has, and whether he has bank accounts and tax debts?", then I am actually not allowed to tell him that. If I ask a police officer: "Could you check whether this taxpayer has antecedents, or whether there are police reports about him?", then he cannot actually tell me, because the police officer also has the obligation of confidentiality. But yeah, if we apply it that strictly and formally, we cannot work integrally. So every Tuesday morning, we have an informal meeting, where all sorts of things are discussed. (...) The police for instance saw someone driving in an expensive car, of whom they know he doesn't work. Then we are asked what we know about that: "Can you check what he earns, what he declares, whether he has savings, whether he can explain that he's driving that expensive car?". Well, then we are going to look in our systems. That's how we share information. Then we're going to discuss what we're going to do with it: are we going to get him under criminal law, (...) or under fiscal law? When the decision is made, we have to share the information formally. Well then the formal requests have to be made, (...) because in 99 of 100 cases a lawyer is involved.'

Chapter 3

The respondent refers to this information exchange as an informal meeting without which integral cooperation would be impossible. Only after choosing the course of action, the formal route is followed. The respondent acknowledges that only the formal side of the story is presented to the outer world:

‘You really have to have your case covered, because otherwise... if you are in front of the judge, and he asks: “Inspector, how did you get these police reports?” “Yeah, I got these at the kitchen table.” Well yeah, then the judge will say it is illegally obtained evidence. Then it’s done, then you’ll never be vindicated.’

The informal meeting makes things possible that are not possible if rules were strictly followed. The participants justify this practice by referring to the common objective of combating crime, for which the ‘take away’ team carries joint responsibility. Social ties with other involved actors also facilitate the obtainment of permits to use information from other organizations:

‘We actually always get the permission, (...), because the prosecutor is [name] from the parquet in [city in Brabant], who also knows us, and also regularly joins our kitchen table. So he knows when we come with this request, then something is the matter.’

The formal request is thus seen as obligatory, rather than necessary to protect citizen-clients. The respondent seems to imply that the prosecutor grants these permits based on strong ties with the team; because he personally knows the members and even joins the informal meetings, it is suggested, the prosecutor can rely on the judgment of the team. Hence, strong social ties between members, as well as the common objective to combat crime, are used as justifications for these ‘informal’ practices.

Multidisciplinary team inspecting separate regulations: Adapting to the team

Multidisciplinary team cooperation between food inspection, police and labor inspectorate is, in the Belgian case, not based on shared discretion. In joint inspections, different agencies enforce different regulations; inspectors are thus formally independent. However, their decisions as part of such teams are oftentimes informally influenced by colleagues’ decisions. In one case, a labor inspector treated a recidivist employer leniently, because a food safety inspector was impressed with his efforts to meet food safety regulations, even though social fraud recidivism is in other cases dealt with by making a report to law enforcement (consistent with agency’s policy). It is striking that the labor inspector emphasizes the employer’s goodwill based on efforts in compliance to rules that her inspectorate is not even responsible for, while the employer fails to correct social fraud, which is considered a very serious violation.

In another case, the police officer in the team had suggested to inspect a new catering establishment owned by one of his acquaintances, because he was sure they were compliant to all regulations. However, many serious violations were detected, putting two labor inspectors in the team in the dilemma whether or not to file a report to law enforcement. They decided not to do so, because they did not want to offend their colleague (the police officer). Instead they made a deal with the employer to give a warning if he would register his employees retrospectively.

‘[Name of police officer] said: “Let’s inspect that establishment.” He knew the owner and was sure that everything was perfectly in order. We [labor inspectors in the team] said: “Are you sure [name of police officer]? Do you really want to do this?” When we entered, we saw them smoking in the establishment [which is a food safety violation]. And no-one of the employees was registered. No-one no-one. He did not speak again all evening. I told him: “We will not make a report.” I even called [name of food safety inspector who was present] that we would not make a report, even though we had found six unregistered workers, because the employer was acquainted with [name of police officer]. I would offer the employer to register them retrospectively and not file a report. But then, the employer had only registered his employees for two hours in the whole month, while the inspection itself had already lasted for more than two hours. So then we [labor inspectors in the team] decided to make a report anyway.’

Although this case ended in the labor inspectors making a report to law enforcement, which would, given the agency’s policy, have been the obvious decision in the first place, this case shows how strong ties and social pressure within interdisciplinary teams can affect the kind of decision being made. Because the same individuals are over a longer period of time responsible for a specific district, the risk indeed exists that inspectors or police officers inspect establishments owned by friends of acquaintances. Dutch tax officials, to the contrary, are prohibited to inspect acquaintances and inspectees living in the same town or area as the inspector is residing.

The experience of social ties by officials, thus, affects decision making about specific cases. In both of these examples, the official felt obligated to adapt his or her decision to a colleague, because of an experience of a social bond with that same colleague. What is notable in these examples, is that the respondents chose for a more lenient treatment than they would have done if they were to decide about this case alone.

Table 3.2. Summary of the main empirical findings

Social constellations	Type of social dynamics	Illustrations
Encounter with entrepreneur	Negotiation: outcome dependent on on-the-spot interaction	<u>Negotiation as a game:</u> ‘...because it’s a game, and they also feel if you’re not standing strong. Then they will open their bag of tricks.’ (Dutch case); a ‘poker game’ in which a ‘poker face’ is essential (Belgian case) <u>Negotiate to obtain greater financial benefits, but for whom:</u> ‘[...] or we do it by means of the anonymous tax rate, but that is of course a much higher correction.’ (Dutch case); ‘[...] she will pay social security taxes for two months and these employees received their money’, which is ‘more important than being able to collect a fine.’ (Belgian case)
	Feeling of bond or hostility: evaluation of relationship affects treatment	<u>Experience of relationship:</u> ‘It depends on your relationship, right? If I have a good relationship with you... well, then I’m the first to say that it is correct, and it influences how you treat someone.’ (Dutch case)
Individual officials within informal collegial networks	Consultation: relying on peers to get an ‘intersubjective’ account	<u>Consult colleagues:</u> ‘You get a judgment supported by colleagues.’ (Dutch case)
Uni-disciplinary, interorganizational team with narrow tasks and individual discretion	Social pressure & social ties: different interpretation and implementation of rule depending on team constellation	<u>For the team, a rule is a rule:</u> ‘The agreement in these anti-fraud teams is to adapt a kind of zero tolerance policy.’ versus ‘In regular inspections (...), you can – and I have already done that – be more lenient (...)’ (Belgian case)
Multidisciplinary, interorganizational team with shared discretion	Strong social ties: informal meeting to open up more opportunities to solve a case	<u>Informal meeting at the kitchen table:</u> ‘[...] if we apply it that strictly and formally, then we cannot work integrally’ (Dutch case)
Multidisciplinary, interorganizational team with broad tasks and individual discretion	Social pressure: adapting decision to a colleague due to experience of a bond with that colleague	<u>Appreciation of a client’s effort towards your colleague:</u> labor inspector was more lenient, because the food safety inspector was impressed by efforts the entrepreneur made to improve food safety (Belgian case)
		<u>A favor for a colleague:</u> ‘I even called [food safety inspector] that we would not make a report, even though we had found six unregistered workers, because the employer was acquainted with [police officer].’ (Belgian case)

3.6. Discussion

This study shows how social dynamics in different bureaucratic contexts influence frontline decision making about specific cases. Street-level bureaucracy literature has described that client-official and collegial interactions may influence frontline decisions (Lipsky, 1980; Maynard-Moody & Musheno, 2000). Our results go further by pointing at the importance of the course of an interaction with citizens, depending on negotiation and bargaining, and thus how both parties stand their ground, in determining the final decision. As for collegial interaction, the literature has mainly focused on the supporting and advisory role of coworkers (e.g. Hatmaker et al., 2011; Van Kleef, 2016). In line with the professionalization and socialization literature, our results confirm that street-level bureaucrats consult colleagues to increase consistency of decision making. These streams of literature, however, cannot explain the more subtle influence of collegial interaction on frontline decision making shown in our results, for example that the mere presence of other professionals alters not only the meaning of rules, but also how clients are addressed and how inspectors perceive and use their discretion. Street-level bureaucracy research should thus broaden its scope by including the influence of these subtle social dynamics on frontline decision making and interactive effects between client-official and various types of collegial interaction.

These results also raise important questions for street-level bureaucracy scholars. First, to what extent does the notion of individual discretion still hold, in light of broader social dynamics that impact decision making? Discretion assumes a degree of freedom and autonomy that is left open by rules and laws, mostly given to semi-professionals with the necessary expertise and skills to handle such autonomy. Ad hoc social dynamics are, however, left out of the equation, while they can hamper autonomy, and lead to unplanned particularism and subjectivism. This study suggests that the notion of individual discretion does not reflect the real-life practice of frontline decision making where decisions are not only the result of thoughtful considerations, but are also impacted by what happens in interactions, such as social ties, social pressure and negotiations.

Second, how do social dynamics within official-citizen and collegial interactions, which are actively encouraged by new responsive policies, relate to fair decision making? Maynard-Moody & Musheno (2003) show how frontline workers strive for making fair decisions by relying on their own moral standards. Fairness is a value that is highly important to these workers. Our study has shown that officials sometimes feel that social dynamics, such as strong social ties, get in the way of fair decision making. When the officials in our examples made different decisions because they felt pressured by colleagues, they valued maintaining a good relationship with the respective colleague higher than the decision they would deem most appropriate or fair. The social character of decision making, thus, could bring other values into the decision making process; officials not only look at case-related aspects in making decisions, but also take into consideration (what certain decisions would mean for) relations with colleagues.

This raises a third question: to whom do street-level bureaucrats feel accountable (if indeed they do so)? Hupe and Hill (2007, p. 290) explain that street-level bureaucrats are 'confronted with multiple demands for accountable behavior', stemming from a multidimensional web of vertical and horizontal relations, that 'may produce tensions'. Whereas balancing between these conflicting demands could in some cases resolve such tensions, the inspectors in this study seem to rather switch between fora of accountability. In other words, on a case by case basis they choose (albeit sometimes implicitly) to whom they feel accountable. Depending on the social context, they feel held accountable by their peers (professional accountability, by for example intersubjective judgments), their clients (participatory accountability, by for example negoti-

ation with clients) or the law (public-administrative accountability, by for instance strict rule enforcement). More importantly, the direction of decisions at the frontline are sometimes the result of unplanned and non-deliberative social dynamics between officials and citizens (negotiations and bargaining) and colleagues (social pressure, strong social ties). As such, decisions are socially constructed, which raises important questions as to whether individual officials – even when formally having individual discretion – can be held fully accountable for the decisions they make.

These findings have practical implications. Prior research has shown that ‘procedural constraints on the exercise of discretion’ (Loyens & Maesschalck, 2010, p. 68), could decrease the risk of inconsistent decision making, but that this, at the same time, could hamper frontline officers to be ‘responsive to individual need’ (Evans, 2016, p. 13) and decrease their motivation. Our study shows that officials sometimes rely on their informal collegial network to increase consistency, particularly in ambiguous situations. We believe more can be gained from strengthening openness among street-level bureaucrats about dilemmas they experience in dealing with discretion, with the aim to find common solutions.

3.7. Conclusion and recommendations for future research

By relying on in-depth interviews and participant observation, this study has studied how social dynamics in official-citizen interactions and in collegial interactions may affect case-related decision making. Our study indicates that social dynamics could play a determining role in frontline decision making in several ways. In negotiating with clients, final decisions are not only dependent on officials’ thoughtful considerations, but also on both parties’ on-the-spot persuasion and negotiation skills. The unfolding of an official-citizen interaction thus plays an important role in case-related decision making. This study has furthermore found that officials try to make their decisions less subjective by relying on their colleagues’ judgments. Besides deliberate consultation, this study has found that social ties with colleagues could also more implicitly affect decision making, leading to a different decision than the individual official would or could have taken. Sometimes this is experienced as positive, as is the case with the informal ‘kitchen table’ meetings, where strong social ties open up opportunities to better solve a case. Sometimes this is experienced as problematic, when for instance social pressure leads to a decision that runs counter to officials’ own idea of what is the appropriate decision.

The main disadvantage of this study is the use of independent case studies with no prior aim to compare. At the same time, using these particular case studies, in different countries and sectors, offered the opportunity to find overall patterns of social dynamics in street-level bureaucracy. Future research on street-level bureaucracy should more often apply a comparative case design to study the impact of different contexts on frontline decision making. Our study could be further extended by examining how frontline officials balance the different values, including those stemming from social interactions, in case-related decision making. Future street-level bureaucracy research could also develop and use research designs that allow better for studying the role of social dynamics in frontline decision making. Participant observation allows for studying how social interactions – and also the more subtle social dynamics such as social pressure – play out in practice, whereas narrative analyses of transcripts of interactions would enable studying what policies, rules and values are evoked and reshaped in real-time interactions (e.g. Bruhn & Ekström, 2017). Social network analyses could be performed to study the influence of frontline officials’ informal network structure and composition on for instance officials’ willingness to reach ‘intersubjective’ accounts.

Chapter 4

Double standards in frontline decision making A theoretical and empirical exploration

Abstract

Drawing on status characteristics and double standards theory, this study explores how social categories may affect the standards tax officials use in evaluating citizen-clients' trustworthiness, leading to differential evaluation. Whereas the street-level bureaucracy literature mainly focuses on the direct effect of social categories on officials' judgments, this study shows how stereotyping in the public encounter could be much subtler and more pervasive than is hitherto studied. Based on semi-structured interviews containing forty stories of tax officials who inspect entrepreneurs' tax returns, this study suggests that similar signals may indeed be interpreted differently for different social groups.

Raaphorst, N., & Groeneveld, S. (*Under review*). *Double standards in frontline decision making: A theoretical and empirical exploration*. Revised and resubmitted to an international peer-reviewed journal.

4.1. Introduction

The decision of 'who gets what and how' is not only made by public officials who are formally endowed with the task to formulate policies, but is also made by the officials at the frontlines of interaction between the state and citizens (Lipsky, 1980; Maynard-Moody & Musheno, 2000, 2003; Hupe & Hill, 2007). Whereas street-level workers have the least formal authority, they have discretion to apply laws, rules and procedures to specific cases. They have discretion 'because the nature of provision calls for human judgment that cannot be programmed and for which machines cannot substitute' (Lipsky, 1980, p. 161). Hence, it is held that street-level bureaucrats' decision making is partly dependent on the relationship they have with citizens, guided by their own cultural beliefs about what is fair and normal (Maynard-Moody & Musheno, 2000, 2003; Harrits & Møller, 2014).

Whereas 'routine cases' facing public organizations are nowadays treated in a standardized way, and handled by automatic computer systems (e.g. Bovens & Zouridis, 2002), the difficult cases cannot be pre-programmed and call for human judgment. In fact, horizontal steering models within public organizations, promoting notions as trust and commitment between frontline officials and citizens (e.g. Yang, 2005; Peters, 2004; Stivers, 1994), even encourage frontline officials' own judgment. Within such governance models, discretion is not only seen as necessary, but also as essential in enabling officials to pursue compliance, responsiveness, and trust (Bartels, 2013). Trust, as a reciprocal notion, shifts the attention from predetermined rules and procedures to aspects of the interaction (Yang, 2005; Stivers, 1994). As such, a tension exists between the organization and the individual street-level bureaucrat; the former encouraging consistency using digital systems and managerial control, the latter opposing this control in the need to be responsive to specific cases (Rutz et al., 2015). This increased street-level discretion thus constitutes a source of uncertainty for managers and citizens, but also for street-level officials who experience ambiguities and dilemmas in discretionary decision making (see findings of chapter 2). The question, then, is how frontline officials working with this leeway come to evaluate citizen-clients as trustworthy or untrustworthy.

Since street-level workers only have little time and information, it has been argued that they look for certain cues or signals to categorize citizen-clients (Mennerick, 1974). Whereas some signals are pre-determined by the organization for which an official works, other signals bureaucrats look for are indeterminate and dependent on the interaction with the client. It is held that in uncertain situations social typologies offer bureaucrats strategic information that they would otherwise not have (Mennerick, 1974). These social typologies 'allow both service workers and clients to fill in the gap between merely knowing the other's formal status and being acquainted intimately with him' (Mennerick, 1974, p. 398-99). The street-level bureaucracy literature commonly focuses on how certain characteristics of citizen-clients, such as their attitude, behavior, ethnicity or socioeconomic status, directly affects officials' categorization judgment (e.g. Maynard-Moody & Musheno, 2003; Harrits & Møller, 2014). As such, these attributes serve as direct shortcuts to citizen-clients' supposed identities.

However, sociological research on hiring decisions of employers relying on the sociological status characteristics theory (Ridgeway, 1991; Foschi, 2000) shows how social typologies not only serve as shortcuts to unobserved characteristics, but also as 'lenses' through which other attributes of the person are assessed. This means that social typologies could not only directly influence officials' categorization judgment, but also indirectly by unequally influencing the interpretation of other signals relevant for the judgment at hand. How social typologies affect the interpretation of other signals or cues and, hence, affect bureaucrats' categorization judgment is however an underexplored theme within research on public encounters. Such a study is import-

ant, since it would give insight in how broader cultural beliefs find their way in official-citizen interactions and affect frontline decision making. Stereotypes that serve as lenses, moreover, are more pervasive than stereotypes that serve as shortcuts, since the former involve stereotypical interpretations of a broader set of aspects.

This study focuses on frontline inspectors working for the Dutch tax authorities who over the years have been encouraged by new models of governance and management to make their own judgment regarding a client's trustworthiness and include this judgment in their decision making. Enforcement of tax laws at the frontline, thus, has become much less clear-cut and predetermined. This paper offers a first exploration of the theoretical mechanisms stemming from the double standards theory in a frontline context. Its aim is to explore how expectations associated with group characteristics may lead to a differential evaluation of similar situations. By analyzing eleven semi-structured interviews with tax officials who inspect the acceptability of entrepreneurs' tax returns, this study poses the following research question: how does citizen-clients' belonging to social groups affect officials' interpretation of signals? In order to examine this, this study first explores more broadly what signals tax officials look at to categorize citizen-clients, and how officials come to interpret certain attributes as signals. For this reason, the first part of the reported findings is more inductive, which allows us to examine what kind of social typologies are used to assess trustworthiness. The second part is more focused and explores how similar signals may be evaluated differently depending on entrepreneurs' supposed belonging to a social group.

In what follows we will discuss two streams of literature that have a different (but not mutually exclusive) take on the role of social typologies in categorization judgments. The first and predominant view in the street-level bureaucracy literature perceives of social typologies as shortcuts to people's supposed identities. The second stream of literature, aligning with the sociological status characteristics theory (Ridgeway, 1991, 2014), studies how social typologies work as interpretive frameworks, also affecting the interpretation of other attributes. After discussing these theories and the specific expectations arising from this, we will discuss our methods and present our findings. In the last section we will discuss our findings and offer recommendations for future research.

4.2. Review of literature

Signaling theory: Social typologies as shortcuts

Street-level bureaucracy scholars commonly acknowledge official-citizen interactions are pervaded by a deep uncertainty (Maynard-Moody & Musheno, 2003; Wagenaar, 2004). Research shows that bureaucrats are 'aware of how little they "really know" about their subjects' (Krauel-Tovi, 2012, p. 382). When street-level workers need to make quick judgments, they tend to reinforce stereotyped or stigmatized identities (Andersen & Guul, 2016; Lipsky, 1980; Maynard-Moody & Musheno, 2003). Studies pointing to the role of stereotypes or shortcuts in officials' categorization judgments (e.g. Gambetta & Hamill, 2005; Harrits & Møller, 2014; Lipsky, 1980; Maynard-Moody & Musheno, 2003; Protas, 1979), implicitly build upon the idea that officials face an information gap which is reduced by looking for observable characteristics. In line with signaling theory (e.g. Spence, 1973; Stiglitz, 1975), it is believed that these observable attributes are linked to unobservable properties (such as trustworthiness), and are bases to make inferences about the latter. Research shows how street-level bureaucrats not only look at more fixed aspects set by the organization to for instance determine a citizen-client's

eligibility for a service, but also at that person's demeanour in the interaction (Maynard-Moody & Musheno, 2003; Nielsen, 2007) and his/her belonging to a social group such as ethnicity (Lipsky, 1980; Epp et al., 2014) or social class (Dubois, 2010). Dubois' study (2010) shows how welfare agents screen applicants on social indicators; agents claim to know whether applicants come from a high or low social class by looking at their behaviour, dress, and use of language (p. 98). The same study shows how welfare agents sometimes verbalize their expectations based on racial stereotypes that mostly turns out to be negative for the minority group.

Following from the latter, we could argue that stereotyping is seen as making inferences about unknown attributes based on people's belonging to a certain social group. Since someone's trustworthiness is not readily observable, we expect officials in this study to use social typologies as shortcuts.

Status characteristics and double standards theory: Social typologies as frames

Whereas research on social typologies as shortcuts points to the signaling function these typologies have, conveying information about unobservable characteristics, the sociological status characteristics and double standards theory shows how people's belonging to certain social groups could also serve as 'lenses' through which other attributes of the person are assessed (e.g. Foschi, 2000). This implies that signals are interpreted differently depending on the social group persons are perceived to be part of. Studies on status characteristics theory are usually conducted within the context of work relations, such as hiring decisions, or decisions within group tasks, and typically focus on explaining why certain groups in society are privileged in attaining positions and rewards over other groups in society (e.g. Wagner & Berger, 1993). Pointing to the influence of broader cultural beliefs on the interpretation of other attributes, status characteristics theory may teach us how stereotyping in the public encounter could be much subtler and more pervasive than is hitherto assumed and studied.

Berger et al. (1972), citing Simmel (1908), argue you might know an individual from direct experience, but also from the individual's status category. Status characteristics are associated with 'cultural beliefs of greater competence in those with more valued states of the characteristic' (Ridgeway, 1991, p. 368). These characteristics are socially recognized attributes on which people are perceived to differ, such as religion, ethnicity, gender, education, and of which the states are differently evaluated. The different states (e.g. being white or black) of a status characteristic (e.g. ethnicity) are associated with different performance expectations in specific task situations (e.g. applying for a job) (Berger et al., 1972). Possessing a certain performance characteristic (e.g. analytical skills) is assumed to increase the likelihood of achieving success at the task. White job applicants, for example, are typically assumed to be more competent than black job applicants. Since evaluators might not know who is more competent in a specific situation, the theory holds, they look at status characteristics that fill this knowledge gap. Status characteristics do not have to be directly culturally associated with the valued task, which means that there need not to be a broader cultural belief about for instance lower educated persons and cooperation skills, in order for an evaluator to interpret these skills negatively based on the person's low level of education. The theory holds that if the status characteristic has not previously been dissociated from the performance characteristic, and is the only basis for social discrimination, this status characteristic will become relevant in this situation (Berger et al., 1972). Status characteristics, then, are typically strong signals in social interactions among strangers. Moreover, the associated expectations typically have a strong effect on the interaction, in that it affects the opportunities people get, the influence they can exert and the type of evaluations

that are made (Foschi, 2000).

Status characteristics theory not only explains the conditions under which status beliefs could matter in the public encounter, but also explains the mechanisms by which status characteristics come to affect the standards officials use to evaluate citizen-clients' trustworthiness. Double standards theory, which is an extension of status characteristics theory, holds that status characteristics not only directly influence people's evaluations, but also affect the standards they use to evaluate other people, leading to unequal evaluations of similar situations. Table 4.1 shows the conceptualizations of the main concepts.

Table 4.1. Conceptualization of main concepts

Signals	Observable attributes that are associated with unobserved properties (for example, level of education associated with competence). The observable attributes could vary from fairly fixed ones (for example, skin color; place of residence) to more easily manipulable ones (for example, level of education; behavior in interaction).
Status characteristics	Socially recognized attributes on which people are perceived to differ, such as ethnicity, gender, education, and of which the states are differently evaluated. Low status groups in society (for example, lower educated; women; blacks) are typically associated with lower performance expectations than high status groups in society.
Double standards	The use of differential criteria to evaluate similar situations

The double standards theory predicts that for members of different status groups, differential standards are used to interpret the same evidence of a given attribute (e.g. competence or morality) (Foschi, 2000). To be more precise, the theory posits that a more lenient standard is applied to a higher status person than to a lower status person exhibiting the same level of evidence of the respective attribute of interest. As such, the use of double standards reinforces the existing status order, because it is harder for a lower status person to be assigned the higher valued attribute (e.g. competence or morality). The research tradition commonly focuses on double standards of competence. However, the mechanism of double standards is also examined in other fields, such as in the inference of attractiveness and overall worth (Deusch et al., 1986), athleticism (Biernat & Manis, 1994), height, weight and income (Biernat et al., 1991) and just earnings (Jasso & Webster, 1997). Also in these studies, status characteristics such as gender and race are the basis for double standards, which are stricter for the lower status persons (Foschi, 2000). In our study, we focus on double standards of competence and intentions. In the following we will illustrate the double standards mechanism by discussing double standards of competence.

The double standards theory distinguishes standards of competence and of incompetence (Foschi, 2000). The double standard for competence works as follows in the context of the public encounter. As performance expectations for low status citizen-clients are lower than those for high status clients, a high performance of a low status citizen-client will be inconsistent with the expectations for lower status citizen-clients. Since the high performance does not accord with the low expectations, officials will need stronger evidence that this performance was indeed the result of competence. As a result, standards will be stricter for lower status citizen-clients. The opposite also holds: as high performance is consistent with performance expectations for high status citizen-clients, the consistency between observation and expectation will lead to loosening the criteria (Correll & Benard, 2006). The official will not need to inquire more deeply,

because citizen-clients' ability could already be inferred from his/her status category. This would for instance imply that street-level inspectors scrutinize a low status citizen-client more thoroughly than a high status citizen-client, when performing equally well.

The mechanism of double standards of *in*competence also leads to a stricter evaluation of lower status groups. Low performance is not consistent with the expectations people have of higher status groups, which lead them to look for more evidence of this incompetence, whereas this is not the case for lower status groups, where low performance is more easily accepted. In our study, we scrutinize how the same signals may be interpreted differently for people from different status groups. Table 4.2 provides a short overview of the double standards mechanism for both low and high performance.

Table 4.2. Double standards mechanism and resulting strictness of treatment

Double standards of incompetence	Low status group	Low apparent competence is congruent with low expectations of low status groups > accepting the negative account (strict)
	High status group	Low apparent competence is incongruent with high expectations of higher status groups > looking for more evidence (lenient)
Double standards of competence	Low status group	High apparent competence is incongruent with low expectations of low status groups > looking for more evidence (strict)
	High status group	High apparent competence is congruent with high expectations of high status groups > accepting the positive account (lenient)

4.3. Research site: The Dutch tax authorities

Our study focuses on Dutch frontline tax officials who inspect the tax returns of small to medium sized businesses. This case was selected on theoretical grounds: the selected tax officials work with a policy that gives them more room for interpretation, and allows them to rely on their own professional judgment for decision making. These tax officials, therefore, represent an apt case to study the possibility and manifestation of double standards.

Tax officials' main task is to evaluate whether entrepreneurs' accounting records are acceptable. They visit entrepreneurs and their business sites to assess whether this is the case. Under the heading of the so-called 'horizontal supervision' approach, the Dutch tax administration has moved from a vertical command and control approach to responsive and collaborative regulation and enforcement (Gribnau, 2007). The official-client interface, then, is not seen as merely an administrative, neutral process necessary to implement policies and in which public officials should be strictly regulated, but as an essential aspect of cooperative and responsive regulation: 'the treatment of taxpayers is based on mutual trust and reciprocity to which good communication is crucial' (Gribnau, 2007: 325). Within the Dutch tax authorities, frontline tax officials are encouraged to assess the acceptability of tax returns, rather than their mere correctness. This means that officials audit entrepreneurs' bookkeeping records not with the idea to correct every gap they find, but with the adage that 'good is good enough' (Belastingdienst, 2016). The standards officials use, are less strict, which makes law enforcement less predetermined and more dependent on the specific context encountered.

In determining the acceptability of the entrepreneurs' accounting records, tax officials then

look at whether entrepreneurs are trustworthy. Based on a first round of interviews with tax officials for a prior study we conclude that tax officials generally look at two aspects of entrepreneurs' trustworthiness – intentions and competences – to assess the truthfulness of what is presented to them, and to assess whether possible gaps between their primary administration and bookkeeping records were caused by mere incompetence or bad intentions. These assessments, in turn, influence officials' decision to more critically audit the enterprise, their willingness to reach a settlement agreement and the height of the possible fine. This aligns with dominant conceptualizations of trustworthiness or sources of trust that generally distinguish two dimensions: goodwill and competence (Barber, 1983; Das & Teng, 2004; Nooteboom, 1996). Nooteboom suggests that trust 'may concern a partner's ability to perform according to agreements (competence trust), or his intentions to do so (goodwill trust)' (1996, p. 990, emphasis original). In our study, competence refers to entrepreneurs' abilities and skills to hold proper accounting records, and thus, to abide by the law. Moreover, assessments of citizen-clients' intentions in general are explored, which could both cover intentions towards the tax administration or tax official as intentions 'in doing business'. This study, thus, examines tax officials' use of double standards in the inference of both competence and intentions.

4.4. Respondent selection

In order to answer the research question, we looked for tax officials who conduct audits and who are given the task of carrying out the 'horizontal inspection' policy. For this reason, we focused on frontline tax officials auditing the tax returns of small businesses, where this policy has been adopted since 2005. We focused on tax officials who inspect tax returns of small and medium sized companies and thereby have direct interactions with citizen-clients, either with or without their accountants present⁶. Since frontline tax officials' contact details were not readily available, we emailed managers. Several managers reacted positively to our request to interview a small number of frontline tax officials. For this study, eleven frontline tax officials from three different offices in the Netherlands were interviewed. Four respondents are female and seven are male. Two respondents have been in service for just over 35 years, one for 24 years, four for 9 years, and four for less than 5 years. The interviews lasted between 68 and 147 minutes (95 minutes on average).

An important part of officials' work is assessing the gaps and errors in citizen-clients' tax returns. They are guided by questions such as: what caused these gaps or errors, what should be the height of the tax correction and should the entrepreneur get a fine? In order to assess the acceptability, officials examine entrepreneurs' bookkeeping records, primary records and operational processes. As part of the acceptability judgment and to assess whether the case merely involves faults, negligence or even fraud, officials look at entrepreneurs' intentions and competences. When present, officials also assess the trustworthiness of entrepreneurs' accountants. The assessment of only the accountants' way of working is however never considered sufficient, because it is the entrepreneur who has insight in the operational processes and primary bookkeeping records. The assumption is that entrepreneurs could still withhold tax money. In the end, the entrepreneur is held responsible for the status of the bookkeeping records and the tax returns. The assessment of entrepreneurs is therefore seen as essential in their evaluation. Tax officials finalize their audit in a report in which they substantiate the decisions on the correction and possible fine, which is sent to the 'audit manager' for a last check before it is sent to the

6. Some entrepreneurs have accountants to assist them in filling out their tax returns.

citizen-client. Whereas one tax official generally has the final responsibility to close the case and hence, make a decision, officials often consult colleagues to reach a decision.

4.5. Methods

For the purpose of this study semi-structured interviews were conducted focused on how tax officials 'get to know' the entrepreneur and his/her enterprise. The interview consisted of two successive parts. The first part was inductive and inquired into the sequential process of an audit; what does a tax official already know in the beginning of an audit and how, and what does the tax official not know? What does the official know after the preparatory phase, when there has been no face-to-face contact yet? How does the official know? And what does the official not know? The same questions were asked with regards to the introductory meeting when the official meets the entrepreneur and the latter's possible accountant. This inductive part enabled us to analyze how officials build a picture of entrepreneurs and their enterprises during the entire process of an audit. This allowed us to explore how social typologies may be used to assess trustworthiness.

The second part of the interview was more structured. A two-by-two scheme was presented to the respondents, with two 'dimensions' of trustworthiness: competence and intentions (see Appendix 2). First, respondents were asked whether these dimensions indeed match their working practices, and whether something was missing from it. All respondents recognized these dimensions as being an important part of 'getting to know' an entrepreneur and his/her accounting. Then, the respondents were asked whether they could recall situations they experienced where they had a strong feeling or idea that an entrepreneur had bad intentions and were incompetent (1), had bad intentions, but were competent (2), had good intentions, but were incompetent (3), had good intentions, and were competent (4). They were specifically asked how they came to this evaluation. Some respondents could not tell a story about each instance, simply because they never experienced it. Sometimes, respondents told more than one story about an instance. In total, forty stories were analyzed. Based on experience, respondents sometimes gave examples about 'types' of entrepreneurs and the latter's behavior. These examples were coded as well.

Both parts of the interviews were analyzed guided by the question what signals and cues frontline officials look at to evaluate citizen-clients' trustworthiness. We considered something to be a signal when respondents interpreted certain fiscal numbers (such as low turnover, or high expense claims) or other observable characteristics as associated with looked-for unobservable characteristics, which are, in this study, intentions and competences. Moreover, we looked at respondents' reasoning about these signals in order to know more about how officials come to interpret certain attributes as signals.

Based on our theoretical discussion, we distinguished two different sensitizing concepts: 'social typologies' and 'double standards'. In the analysis, we distinguished between social typologies that have particular meaning for the tax authorities (such as 'starting enterprises' and 'financial advisors') and social typologies that are shared by society at large (such as 'the higher educated', or the 'foreigner', which can be depicted as status characteristics). We coded 'double standards' when similar signals were interpreted differently for entrepreneurs from different status groups, i.e. groups distinguished by their status characteristics. To scrutinize the latter, we studied how individual respondents may evaluate the same signal differently for different 'types' of people.

We inductively found five main empirical clusters: characteristics of bookkeeping; prior knowledge; interaction; social typologies; characteristics of client's home or private situation.

During the coding process, the clusters were filled with more detailed, distinctive themes and references to the transcripts. Constant comparison between codes allowed us to adjust or merge codes. Overall, the clusters ‘characteristics of bookkeeping’ and ‘interaction’ contained most of the references.

4.6. Findings

Assessing the citizen-client: signals of trustworthiness and untrustworthiness

To evaluate citizen-clients’ trustworthiness, frontline tax officials look at a variegated range of signals and cues. These signals have different natures and stem from different sources. Table 4.3 provides an overview. The table distinguishes between signals of competence, incompetence, good intentions and bad intentions, and shows the most frequently mentioned signals, or exemplary signals where frequencies are equal. We distinguished different sources of signals that could be seen as being part of the fiscal procedures of auditing on the one hand, and sources of signals that are part of the social interaction with entrepreneurs on the other hand. As part of the auditing strategies, officials look at characteristics of the bookkeeping records, the tax returns, prior knowledge about specific enterprises, and they also take into account the statuses certain professions have within the tax administration.

Our analysis shows that officials not only look at signals related to entrepreneurs’ bookkeeping and tax returns, but also at their demeanor in the interaction, their belonging to social groups in society, and their home situations. The source ‘demeanor’ involves both entrepreneurs’ non-verbal behavior in the interaction as the verbal statements they make. Entrepreneurs’ belonging to a social group is another source of signals officials look at. Moreover, officials sometimes mention looking at people’s home situations to make inferences about for instance how much money someone owns, or about the possible cause for gaps or faults found in the bookkeeping records. The signals do generally not have a conclusive meaning on their own. It is the collection of signals, gathered through the entire auditing process that is decisive for the trustworthiness judgment. This suggests that signals are interpreted differently based on the whole picture that has been formed up to that moment.

Table 4.3. Sources of signals (leftmost column) and examples of kinds of signals looked at to assess trustworthiness and untrustworthiness

	Competence	Incompetence	Good intentions	Bad intentions
Auditing procedures				
Bookkeeping & tax returns	Neatly organized	Messy or no records	Having arranged backup from advisor	Bad tax filing behavior (too late)
	Segregation of duties	Wrongly interpreting rules		Claiming tax but not paying tax
Prior knowledge	Major, reputable accounting firms	Advisor who is known for wrongly keeping records	Major, reputable accounting firms	Entrepreneur is known for fraudulent behavior
Status of profession in tax administration	Financial advisors	Starting enterprises	Financial advisors	Automotive sector
			Starting enterprises	Chinese restaurants
Assessment of particularistic attributes				
Demeanor	Cooperative	Rolling eyes	Cooperative, open	Hostile and not constructive
		Not acknowledging weakness	Admitting fault	Making conflicting statements
Belonging to social group	Highly educated	Baker, butcher, bike seller	-	Residents of mobile homes
				Chinese
Home/private situation	-	Sad home situation (disabled child, shabby interior of house)	-	Luxurious home and car, but no money

Respondents noted that the preparatory phase, which is the phase prior to the face-to-face interaction and actual audit, is important in forming an initial image of the enterprise and how it is doing in fiscal terms. There are a lot of internal systems and sources of information (both public and not public) tax officials can consult to gather signals. The findings of the preparation feed into officials' considerations regarding how to go about the audit. Only few signals in this phase are conclusive about whether something is wrong, and sometimes also about someone's intentions. Negative prior knowledge about an entrepreneur or financial advisor, such as corrections within a prior audit, generally serves as a strong signal that it is likely to be wrong again. Most signals are however points of attention, roughly distinguishing between 'it looks good' or it 'does not look good', offering possible starting points for the further audit. One example is entrepreneurs who claim a lot of costs, but do not have revenue, which means that they claim VAT, but do not pay taxes. This might indicate that entrepreneurs withheld tax money.

Officials also form expectations based on generalized assumptions about certain professions. One example is the assumption that within the automotive sector there is a lot of fraud, leading officials to expect no good when having to audit such an enterprise. Within sectors where a lot of cash money circulates, such as taxis or hospitality businesses, the expectation is that a lot of tax money is being withheld, since 'cash money is burning in the hands' (Respondent 5).

Although not offering a conclusive picture, the signals found in the preparatory phase are critical to how the further audit process is conducted. That is, whether a citizen-client is inspected more critically, or only minimally.

All respondents emphasize that the actual interaction with the entrepreneur gives stronger signals as to the latter's intentions and competences than the preparation. One respondent argues: 'only when I have been there, assessed the records, have seen and spoken to the entrepreneur, have received a guided tour, and the impressions I have of this, I can write down something meaningful' (Respondent 5). Another respondent even admits that he, against the rules, skips the preparatory phase, because he thinks it is better to not have prior expectations about the entrepreneur which, according to him, creates a 'tunnel vision'. He prefers talking with entrepreneurs with an open view, which he deems 'more important than the numerical substantiation of the preparation' (Respondent 11). Respondents typically assess whether entrepreneurs offer good explanations for the points of attention found in the preparatory phase. Generally speaking, officials thus perceive the signals in the interaction as more important for their evaluation than signals they pick up in the preparation.

In view of this way of working, it is not surprising that tax officials also go beyond the book-keeping and tax returns, to look at citizen-clients' demeanor in the interaction to assess their trustworthiness. Respondents' stories and examples bear witness to an implicit procedure they rely upon to gather signals. To assess whether entrepreneurs' records are acceptable, officials determine the so-called 'situation as expected' and 'the situation as it is'. The former is the situation that can be expected based on the norms, which are typically based on averages of figures of comparable enterprises, and on rules and laws. Deviations from the norms are points of attention that need to be further examined; what explanations do entrepreneurs have for the discrepancies?

This assessment of the actual situation as compared to how it should be, does not only occur in fiscal terms, but also in social terms. Officials look at how entrepreneurs behave in the interaction in terms of what they deem 'normal behavior and attitudes'. This means that they look at deviations from normality in the broadest sense of the word. Citizen-clients who for instance shiver or blink their eyes a lot are looked at with suspicion, since it might indicate they are nervous about something (Respondent 4). Offering a cup of coffee (respondent 3), talking very fast (respondent 11), sweating (respondent 5) are other examples of non-verbal cues that are looked at to make inferences about someone's likely intentions. Officials furthermore look at what entrepreneurs say, to assess whether their statements correspond with each other, but also with observations they have made in the preparation or during the guided tour through the enterprise. Entrepreneurs' home situations are hence sometimes scrutinized to look for possible explanations of deviations from norms (i.e. suffering financial hardship), but also to assess whether there are discrepancies (i.e. declaring you don't have money, but possessing expensive things). The assumption is that these discrepancies are not comprehensible in light of the normal (honest and competent) entrepreneur who would be open, cooperative, consistent, etcetera. This relates to the concept of normality as used by Møller & Harrits (2013), who define it as an ordering mechanism, 'identifying which aspects of social life are being dashed as risky and which aspects are seen as indicators of normal behavior' (p. 158). Our analysis shows that normality assumptions are not only driving the audit of the records, but also officials' assessment of entrepreneurs' demeanor in the interaction. Whereas the notion of normality is more explicit in the preparatory phase, since it is mostly based on averages, it is more implicit and based on individual officials' prior experiences within the interaction.

Another procedure officials rely upon to gather signals, is looking at whether entrepreneurs

belong to a certain status group that is shared by society at large. This way of gathering signals is not based on assessing deviations from norms, but based on assessing similarities between known status groups and characteristics of the respective citizen-client one is facing. When grouped into a certain status group, this leads officials to have certain expectations about that person's competences and/or intentions. For instance, people who primarily work with their hands such as bakers, butchers and bike sellers are expected not to be good at doing their records, because they 'like to work with their hands' (Respondent 1). It is assumed by one respondent that Chinese people, to mention another example, like to gamble with money, and go to casinos to gamble with tax money (Respondent 5). The same respondent even argues he thinks that all Chinese people do this, because it is part of their culture. Another respondent argues that 'it is often the simple minded who start a business', and he continues that although these are often 'good people', that it is often the 'lower classified people in society who muddle through, who start a scrap business somewhere'. He associates a specific expectation to these types of citizen-clients: 'then you think, why do you start such a business for God's sake? You know there will be big tax bills, and that he will be ending up at the tax collectors again' (respondent 3).

To sum up, tax officials elicit signals from following the auditing procedures and consulting digital information systems, but also from more particularistic sources such as their encounter with the entrepreneur, where they make inferences based on entrepreneurs' demeanor, belonging to a social group, and their private or home situation. There are generally three ways in which officials come to interpret certain attributes as signals. First, by relying on prior knowledge about individual entrepreneurs, assuming that they are not highly likely to change overnight. Second, by assessing the ideal situation (the norms) with the actual situation, looking at the entrepreneur's bookkeeping, demeanor and statements in the interaction, and her/his home situation. Deviations from the norms indicate something may be wrong. Some of these norms are, thus, part of the fiscal procedures of auditing, others are more particularistic, dependent on officials' own norms of what is normal behavior. Third, by assessing whether citizen-clients belong to a certain social group, based on which officials form expectations about the formers' competences or intentions. Signals do not have a conclusive meaning on their own, but are interpreted in light of other signals that have been collected through the inspection process. Similar signals can thus be interpreted differently depending on the picture that has been formed. In what follows, we examine whether these differences are based on status characteristics.

Exploring double standards mechanisms

To examine our main research question, we studied how officials may evaluate similar signals differently for entrepreneurs from different social groups. Because double standards theory posits that individual evaluators use differential standards to evaluate similar signals, we did this analysis on the level of the individual official. We particularly looked at whether respondents who talked about different social groups, evaluated similar signals differently depending on the expectations they have of those group members. We explored the possible relations between status group characteristics, expectations about those group members, and differential evaluation of signals. We found that the signals 'wrongly kept records' and 'wrongly declared turnover tax' are associated with a low quality of the bookkeeping, but that these signals are interpreted differently for different social groups by three respondents (see Table 4.4). In what follows, we will elaborate on how similar signals are evaluated differently.

Table 4.4. Status characteristics as framework leading to different interpretations

Signal	Status characteristic	Interpretation
Respondent 1		
Wrongly kept records	Residents of mobile homes	Bad intentions, incompetent (they can't and don't want to)
Wrongly kept records	People who work with their hands	Good intentions, but incompetent (they just can't do it)
Respondent 6		
Wrongly kept records	Lower educated / underclass of society	Bad intentions
Wrongly kept records	Higher educated	Good intention gone bad
Respondent 12		
Wrongly declared turnover tax	Shoemaker	Incompetent, not to blame
Wrongly declared turnover tax	Mayor	Competent, to blame

These findings show that the differential standards that are used to evaluate low quality of the bookkeeping are actually double standards used to evaluate entrepreneurs' intentions. Whether these are evaluated as either good or bad is dependent on (the official's perception of) an entrepreneur's belonging to a social group⁷. Respondent 1, to elaborate a bit further, evaluates wrongly kept records as an expression of incompetence for both craftsmen and residents of mobile homes. However, craftsmen are assigned good intentions, whereas the residents of mobile homes are assigned bad intentions. The following interview excerpt shows the reasoning the respondent employs in assigning the craftsperson good intentions.

'Someone can be very good in doing his job, you especially see this with people who work with their hands, who make something. They can be very good craftsmen (...). But they also need to keep their records and they are thinking like, well... I hate that so much, someone else has to do that. And they sometimes keep their records very badly. Then it's not so much that they don't want to, but more because they are too busy with other stuff. A bike repairman, for example, is very busy repairing bikes and then sometimes forgets to write down what he exactly used, you know? And then later on, he thinks, that's gone and that's gone, I probably used that. You know, like that.' (Respondent 1)

The respondent, thus, states that bad bookkeeping records of craftspeople are not necessarily an expression of unwillingness, but the result of being too busy doing manual labor. For residents of mobile homes, to the contrary, the respondent interprets a similar signal as an expression of bad intentions:

7. Whether status groups have a low or high status is dependent upon the reference category or categories a group is compared to. Some status groups that are mentioned by respondents can clearly be depicted as having a low or high status in society, such as the lower and higher educated. Whereas other groups not necessarily have a high status in society, but still have a higher status than other groups.

‘Yeah, it’s almost stigmatizing a bit, but to mention an example: the residents of mobile homes. They never learned anything, they often cannot even write. They cannot write and read, but they also thankfully misuse that, because they also know like “they cannot catch me on that”, so to speak. Yeah, as a matter of fact, they cannot get away with that, because they have to do that, but if nothing is recorded and everything is in cash, yeah, who knows where it has been then. You cannot find anything of that. Then you start looking, like, yeah he does have a nice house, and a big car, yeah then they are often [laughs] smart enough to cover it up a little. And then you know, they must have earned something. But their bank accounts, for example, are always in the red, or almost in the red, so they are like “we can barely pay our groceries.” But from the inside you see the gold is shining. Yeah, but you first have to prove it.’ (Respondent 1)

Compared to the expectation about manual workers, this quote shows that a similar signal (bad bookkeeping records) is interpreted differently for the residents of mobile homes. The respondent presupposes that residents of mobile homes have bad intentions and misuse their disability to write and read to commit fraud. This presupposition also leads to a more critical inspection approach when the records are not kept properly by people within this social group, since the respondent starts looking at people’s possessions and their house interior as signals for socioeconomic status. The last sentence of this quote ‘but you first have to prove it’, underlines that this evaluation of bad intentions is primarily based on presuppositions about this particular status group that are extrapolated to particular entrepreneurs. Although both ‘types of’ entrepreneurs show similar evidence regarding the quality of the bookkeeping, this is evaluated with different standards based on who they supposedly are.

Another respondent assigns bad intentions to entrepreneurs who do not keep their records properly, regardless of their level of education. However, the reasoning that is behind this differs. For the lower educated persons who constitute the underclass of society, this respondent presupposes that they are withholding money because of their difficult financial position:

‘What is particularly the case, is that the staff of taxi companies mostly constitute the bottom side of the labor market. People with a low... let’s say, people who didn’t learn a lot, but do know the things they need to know. But uhm, I think fifty percent is a street taxi driver. Which are the people you approach at the station to bring you home. Tips are being given. And if you look at the records, there are no tips included. The entrepreneur [of the bigger taxi company] is not looking at that, and the driver will not record these rides, with which he is stealing from his boss, so to speak.’ (Respondent 6)

It is assumed by this respondent that these ‘types’ of entrepreneurs are more easily inclined to steal from their boss. The respondent even argues later on that when someone has payment arrears, one easily encounters problems, especially when the business is mostly based on cash money. Within the following story, the respondent likewise thinks the entrepreneur has bad intentions. However, the reasoning in this case is different, based on the entrepreneur’s high level of education.

‘He’s on the edge, because he just thinks “I’m not paying”. He’s actually not competent. He doesn’t know how to really fill out the turnover tax return, so to

speak, but he is capable enough to fill it out the way he does. But he's not able to actually keep his records. He doesn't know about that. (...) But he knows how to fill out null in his return. And if he thinks like I can still get 130 euros back because I have the invoice, then I claim 130 euros back. So he's on the edge. He knows what he's doing. (...) I came there, and he appeared to be a MA, in business economics. He first completed his bachelor, and then did business administration at the university. Well, then you can assume that he has a certain level of knowledge. He, thus, should also be able to keep records. He knows what is fiscally possible and what is not. So, uh, in first instance he... may have wanted to do it right. But well, his family was involved, and they probably didn't have good intentions. I did the inspection, and the bad intentions were evidenced by his daily log. Everything fitted for a hundred percent. That is not possible, that is too good.' (Respondent 6)

This respondent associates the entrepreneur's high level of education with good intentions, because he knows how to do his bookkeeping records. The respondent believes that these good intentions were corrupted by his family, who are assigned bad intentions. This presupposition that the entrepreneur was 'on the edge' of bad intentions, in turn, influenced the way his records were interpreted. In fact, the respondent holds that his bookkeeping records looked too good to be true. Whereas both the lower educated entrepreneur and higher educated entrepreneur did not keep their records properly, the interpretation of this signal is different for both cases. For the lower educated entrepreneur who keeps something out of the records, this signal is interpreted as a purposeful act and as 'stealing'. For the higher educated who does the same, this signal does not lead to a total devaluation of his intentions. In fact, his intentions are interpreted as 'on the edge', which came about as the result of an external factor, namely the supposed bad intentions of his family.

Whereas the abovementioned examples are differential interpretations that were not explicitly compared by the respondent (that is, the residents of mobile homes were not compared to the manual workers by the respondent herself), within another example a comparison between status groups and concomitant differential interpretations are openly made.

'A colleague inspected a mayor who had done something wrong with turnover tax. Yeah, then I think, at that level! It was such a big mistake. I thought that if you would have asked a three-year old kid he would know... And then I think he's to blame. It's another story if you're inspecting a shoemaker, for instance. A lot of this is not his expertise.' (Respondent 12)

This example clearly shows how the same signal (wrong turnover tax) is differently interpreted for someone who is considered smart as opposed to someone who works with one's hands and has less expertise.

To sum up, the findings indicate that double standards mechanisms may work in different directions. Lower status entrepreneurs are sometimes evaluated more strictly than higher status entrepreneurs, but our interview data shows it could also work the other way around.

4.7. Conclusion and discussion

To evaluate entrepreneurs' trustworthiness, frontline tax officials look at a variegated range of signals and cues. They do not only look at signals related to entrepreneurs' bookkeeping, but they also look at entrepreneurs' appearance, demeanor, home situations and belonging to social groups to make inferences about the latter's competences and intentions. In accordance with the stream of literature studying social typologies as shortcuts, we thus found that tax officials rely on social typologies of entrepreneurs to form expectations and thereby reduce uncertainty. In addition to this, we explored *how* social typologies may also serve as lenses affecting the interpretation of a range of other signals. In line with double standards theory, two respondents evaluate the social group that has been ascribed a lower status in society (residents of mobile homes and the lower educated) with bad intentions when encountering a bad quality of the bookkeeping, whereas their interpretation for the higher status group (manual workers and higher educated) is more lenient or nuanced. Another respondent, to the contrary, openly declares he has stricter standards for a high status entrepreneur such as a mayor when encountering mistakes than for a lower status entrepreneur such as a shoemaker. The standards tax officials use, thus, may be different depending on who entrepreneurs supposedly are, i.e. to which social group they belong.

Our findings have implications for the branch of literature focusing on street-level officials' use of stereotypes. Whereas the dominant perspective mostly studies how social typologies serve as shortcuts or direct cues (e.g. Andersen & Guul, 2016; Lipsky, 1980), this study shows how stereotyping at the frontline may also work more implicitly by affecting the standards officials use. The double standards theory offers a theoretical mechanism for differences in interpretation of similar evidence. It goes beyond a mere focus on direct cultural beliefs. This means that street-level officials need not necessarily have to employ direct cultural beliefs about specific citizen-clients' trustworthiness (e.g. 'the higher educated have bad intentions in general') in order to make stereotyped judgments that are disadvantaging certain social groups. To be more precise, social typologies sometimes instigate stereotyped judgments when they occur in combination with other signals (e.g. 'bad filing behavior is a signal for bad intentions for the higher educated, but not for the lower educated'). This subtler and more pervasive way of stereotyping calls for research approaches that take into account clusters of signals and focus on how different signals could interact with each other.

When we look at the specific mechanisms underlying the use of double standards in our study, we have seen that it can work in two different directions: the standards can be stricter for the low status entrepreneur and more lenient for the high status entrepreneur, or they can be stricter for the high status entrepreneur and more lenient for the low status entrepreneur. The literature on double standards mostly focuses on the former, since these constitute the core of the theory. The reversed practice, i.e. more lenient standards for the low status person, has been advocated by some as an instrument to change the status quo (Foschi, 2000). However, it also results in inequality, since it reinforces the assumption that the lower status person cannot meet the universalistic standards and therefore has to be treated more leniently (Foschi, 2000). This was also seen in our data, where one respondent believes that the craftsman is not to blame, because of incompetence. In one way or another, the use of double standards tends to reinforce the existing, unequal status order. This exploratory study, thus, found that evaluations of similar situations can differ for citizen-clients from different social groups, and that this can work in different directions. Future research should focus on explaining the individual level differences more systematically, and examine how officials come to develop either double standards or reversed double standards when evaluating citizen-clients.

This study employed semi-structured interviews and, for that reason, had little control over the research context to test whether the same signals are evaluated differently for different social groups. The fact that no direct questions were asked about stereotyping, and the sensitivity around talking about this topic, could be a reason why only few respondents employed stereotyped reasoning. Also, the use of double standards was sometimes implicit, making it hard to probe into it further. By letting respondents evaluate vignettes (hypothetical scenarios) that portray clusters of signals, future research should test the double standard mechanism more systematically and among a larger group of officials, and in different contexts. Such a design also allows for the study of the magnitude of the use of double standards within street-level contexts.

The findings of this study also have implications for new models of governance and management that have come to embrace street-level officials' own interpretation of citizen-clients as essential for decision making. Promoting trust and collaboration between street-level officials and citizen-clients, such models endow street-level officials with substantial discretion to use their own standards for judgment and decision making. This study has shown that, in such a context, street-level officials are prone to make evaluations based on particularistic attributes such as citizen-clients' home situations, but also their presumed belonging to social groups. More decentralized governance models with less clear-cut and predetermined law enforcement, thus, seem to imply a move away from impersonality and equal treatment towards judgments based on aspects of the interaction and citizen-clients' status characteristics. In fact, in the absence of clear-cut and guiding procedures, and vested with the task to evaluate interpersonal notions as trustworthiness and collaboration, officials are given the opportunity to use their own particularistic and sometimes differential standards to evaluate citizen-clients. The evaluations officials make based on double standards have significant consequences, since this practice affects the distribution of opportunities and service allocation and thereby reinforces inequalities existing in society. This raises important questions for public management scholars focused on the question of how public organizations are managed to multiple ends such as consistency and responsiveness (e.g. Piore, 2011; Rutz et al., 2015).

We focused on a particular type of frontline workers, who have a lot of room for interpretation and work with specific types of citizen-clients. Within a context where frontline officials have less discretionary room and more stringent guidelines, there may be fewer opportunities for them to use stereotypes and double standards. Also, the content and use of double standards might be different in other street-level contexts, such as classrooms and social service organizations. As part of their inspections, tax officials are assessing whether entrepreneurs are competent and have good intentions. Social benefit providers might be less concerned with citizen-clients' competence, and more with their intentions. Police officers, to mention another example, might use double standards in assessing citizens' potential criminal intent or behavior. The social typologies that inform the double standards may thus also differ across different street-level contexts, and even across countries. Future research could compare these different contexts, to develop a theoretical framework explaining the use of double standards by frontline workers more broadly. Moreover, we have seen that officials use social categories such as 'residents of mobile homes' or 'crafts people' as lenses to evaluate other signals. These broader social categories tend to intersect with other lower status categories in society, such as lower social class and other cultural backgrounds. Future research could study how intersectionality affects evaluation standards of officials.

Lastly, the findings have implications for the practice of managing street-level bureaucracies. This study draws attention to a subtler way of stereotyping than the research tradition on street-level bureaucrats' use of social stereotypes as shortcuts. Since the use of double standards

Chapter 4

entails evaluations of 'objective criteria' against an invisible standard, and not the direct evaluation of signals, the practice of stereotyping at the frontline may be less visible and harder to detect than is hitherto assumed. Although officials may base their decisions on 'hard evidence', the interpretation of this evidence may differ across groups of citizen-clients. It shifts attention from a sole focus on direct stereotyped expectations (e.g. 'the lower educated are less competent') to indirect stereotypical judgments made against a standard (e.g. 'for a lower educated person, this evidence signals competence'). This is of interest to both public managers and street-level bureaucracy scholars. This first exploration of double standards theory in a street-level context has shown that indirect mechanisms of stereotyping may be of additional explanatory value in research on stereotyping by frontline officials.

Chapter 5

Do tax officials use double standards in evaluating citizen-clients? A policy-capturing study among Dutch frontline tax officials

Abstract

In line with psychological and economic discrimination theories, street-level bureaucracy studies show a direct effect of citizen characteristics on officials' judgments, or show how street-level bureaucrats employ stereotypical reasoning in making decisions. Relying on sociological double standards theory, this study hypothesizes that citizen-clients' status characteristics not only directly influence officials' evaluations, but also indirectly and more pervasively by influencing the interpretation of other signals. By means of a policy-capturing study among Dutch frontline tax officials, this study takes a first step in testing double standards propositions in the context of official-citizen encounters. The findings support only some hypotheses, but indicate that citizen-clients' level of education could serve as moderating contexts affecting the interpretation of cues. The article provides important theoretical and methodological guidelines for future research on stereotyping at the frontline.

Raaphorst, N., Groeneveld, S., & Van de Walle, S. (*Under review*). Do tax officials use double standards in evaluating citizen-clients? A policy-capturing study among Dutch frontline tax officials. *Revised and resubmitted to an international peer-reviewed journal.*

5.1. Introduction

Street-level bureaucrats typically have considerable leeway to make judgments about citizen-clients (Lipsky, 1980). Research on street-level bureaucrats, such as police officers or teachers, has shown how discretionary judgments sometimes overlap with citizens' supposed belonging to certain social groups, such as someone's race (e.g. Epp et al., 2014; Strohshine et al., 2008), social class (Harrits & Møller, 2014), or gender (Johnson & Morgan, 2013). It has been shown that, due to a lack of information, time and other resources, street-level bureaucrats develop shortcuts such as stereotypes to categorize clients (Lipsky, 1980; Protzas, 1979). In situations with only limited information and time pressure, the matching of citizen characteristics to stereotypes gives officials information they would otherwise not have (e.g. Maynard-Moody & Musheno 2003; Gambetta & Hamill, 2005).

Within the public administration literature there is a lack of explanatory studies focusing on how cultural beliefs about social groups play a role in the public encounter, and affect the judgments of frontline officials (but see Andersen & Guul, 2016; Harrits & Møller, 2014; Schram et al., 2009). This is particularly interesting given the fact that frontline officials are encouraged to be flexible and to be responsive to citizens' situations when making decisions. In fact, interpersonal notions as trust and collaboration have come to play an important role in frontline decisions (Yang, 2005). In such contexts, officials have more room for interpretation, and to use their own standards of who is trustworthy and who is not. Therefore, this flexibility paves the way for stereotyped images and double standards to inform judgments.

The sociological status characteristics theory holds that in situations entailing interpersonal task situations, where there is a distinction between 'failure' and 'success', evaluators look at people's status characteristics to evaluate their likely behavior and achievements (Berger et al., 1972). These characteristics are socially recognized attributes on which people are perceived to differ, such as ethnicity, gender, or education, and of which the states are differently evaluated. Status characteristics are associated with 'cultural beliefs of greater competence in those with more valued states of the characteristic' (Ridgeway, 1991, p. 368). As a consequence, it is held, similar situations implying equal competences, are evaluated differently for lower status groups than for higher status groups. By testing the explanatory power of double standards theory using a policy-capturing design, this article sets out to examine how stereotyping at the frontline may be more indirect (i.e. also indirectly leading to unequal judgments) and pervasive (i.e. affecting the interpretation of other signals) than is hitherto studied within public administration research. This study thereby provides a first step in testing the explanatory potential of the sociological double standards theory in a public administration context, and proposes a methodology that could be used to study this overlooked phenomenon.

In what follows, we will discuss previous research on stereotyping in frontline work more broadly. We will subsequently present our theoretical framework, describe the research setting and formulate hypotheses. Then we will describe the used policy-capturing method and discuss our findings.

5.2. Stereotyping at the frontline

The literature on stereotyping at the frontline is diverse and entails different perspectives on stereotyping. Notwithstanding the differences, most of these studies focus on direct ways of stereotyping, i.e. how decisions or evaluations are affected by stereotypes or based on stereotypical reasoning. In the following we will discuss these studies, and show that indirect ways of stereotyping have barely been studied within research on frontline stereotyping.

In line with the economic theory of statistical discrimination, there are studies that assume that the use of stereotypes is based on statistical knowledge or prior experience to reduce uncertainty (e.g. Gambetta & Hamill, 2005; Harris, 1999; Stroshine et al., 2008). Studies show how service workers in general or officials within certain professions explicitly construct suspicious types of clients that are inextricably linked to certain groupings in society. Stroshine et al. (2008) for example, show how police officers find black people driving in dilapidated cars in white neighborhoods suspicious. Gambetta and Hamill (2005) show how taxi drivers find a group of male customers more suspicious than a group of female customers. Due to the sometimes dangerous nature of their work, these workers need to be able to make quick judgments. They rely on such stereotypes to be able to for instance quickly decide whether to arrest someone, or whether to allow someone to enter their taxi. Within these studies, the mechanism of discrimination studied is direct: cues lead people to distinguish between social categories regardless of any other relevant characteristics. Observational studies that point to the stereotypical reasoning employed by frontline workers in reaching decisions (e.g. Maynard-Moody & Musheno, 2003; Dubois, 2010), also fall in this category since they point to how differential evaluations of for instance deservingness overlap with distinctions between social groups.

Within the street-level bureaucracy literature there are only some studies that focus on indirect mechanisms of stereotyping. A study by Harrits and Møller (2014) shows how social workers' tendency to suggest interventions in similar situations is different for low and high class citizens than for middle class citizens. Drawing on the sociological literature on normality and categorization, they find some evidence that the social distance between workers and citizen-clients in interactions implicitly influences their judgments. Moreover, the experimental vignette study by Schram et al. (2009) on case managers' decisions to impose sanctions, shows that black welfare clients are more likely to be punished than white welfare clients when rules have been violated. They test the Racial Classification Model (RCM), a model they developed themselves, to explain how client race affects officials' evaluations of rule violations. The RCM posits that when cues are confirming negative racial stereotypes, this can provide expectancy confirmation, thereby reinforcing negative stereotypes in evaluators' minds. As such, that study also tested and provided evidence for an indirect mechanism of stereotyping.

Apart from these studies, there is little attention within the public administration literature for indirect mechanisms of stereotyping. Our study builds on these studies by testing propositions of the sociological double standards theory to explain how stereotypes may also work as frames affecting officials' interpretation of similar evidence. Just like the RCM, double standards theory holds that negative cues are interpreted stricter for social groups which have a more negative status in society. The double standards theory, however, differs from the RCM in several respects. First, the double standards theory has a broader scope and not only offers explanations for stereotyping based on race, but also on other characteristics, such as gender, social class and educational attainment. Second, whereas the RCM only focuses on 'discrediting markers' or negative cues, double standards theory also offers explanations for how positive cues could be evaluated stricter for lower status groups than for higher status groups. Double standards theory, thus, has a broader applicability and offers explanations for stereotypical evaluations of both stereotype-consistent and stereotype-inconsistent cues.

By means of a policy-capturing study among Dutch frontline tax officials, this current study tests the proposition that tax officials use double standards in inspecting entrepreneurs from different status groups. The previous qualitative study on tax officials reported on in chapter four, suggests that status characteristics, such as ethnicity, level of education and social class, not only serve as direct signals, but may also serve as frames through which other signals are in-

terpreted, indirectly leading to unequal treatment. An analysis of tax officials' stories indicated that similar signals were interpreted differently for entrepreneurs from different status groups, such as the lower versus the higher educated. Since the qualitative study could only explore the double standards mechanism inductively, a more systematic approach to study the effects of status characteristics on officials' interpretations of signals is warranted. A policy-capturing study allows for the study of the effects of status characteristics by using vignettes.

5.3. Theoretical framework

In order to test if and how status characteristics affect the interpretation of other signals, we draw on the sociological status characteristics theory and double standards theory. These theories of status-based discrimination have their origin in the sociology of work, where they have been tested to explain why certain groups are privileged in attaining positions and rewards over other groups in society (e.g. Wagner & Berger, 1993). Status characteristics theory has been depicted by Wagner and Berger (1993) as a program of interrelated theories, aimed at explaining aspects of status-based discrimination in social interaction. In what follows, we will discuss double standards theory (e.g. Foschi, 2000), that extends status characteristics theory by proposing that status characteristics affect the standards people use to evaluate other people.

Double standards theory proposes that status characteristics affect the standards evaluators use to determine other people's ability (Correll & Benard, 2006; Foschi, 2000). The basic assumption is that standards are stricter for lower status groups than for higher status groups (Foschi, 2000). As performance expectations for low status groups are lower than for high status groups, a high performance of a low status actor will be inconsistent with the expectations for lower status actors. As a result, double standards theory holds, standards will be stricter for lower status actors, i.e. they will be more critically scrutinized in this situation. A woman with three children and an outstanding CV, for example, will be looked at with suspicion, because it does not correspond to the lower expectations people have of mothers in the workplace. Employers are inclined, for instance, to look for evidence that disproves the achievements of this person. The opposite also holds: as equally high performance is consistent with performance expectations for high status actors, the consistency between observation and expectation will lead to a more lenient standard (Correll & Benard, 2006). A man with three children and an outstanding CV, in this case, is not inconsistent with the high expectations people usually have of men in the workplace. As a consequence, employers will accept that person's portrayal of his CV more easily, without looking for more evidence of his competence.

Standards of both competence and incompetence can be distinguished. A strict standard for competence requires more evidence than a lenient one, whereas a strict standard for incompetence accepts less evidence of incompetence than a lenient standard (Foschi, 2000). The latter means that cues indicating low competence are more easily accepted for a lower status person than a higher status person, because they are consistent with the low expectations of the lower status group's competence and inconsistent with the high expectations of higher status group's competence. To sum up, the theory holds that both cues signaling low competence and cues signaling high competence are evaluated stricter for lower status groups than for higher status groups.

This chapter examines whether tax officials use double standards in evaluating various cues from entrepreneurs from different social groups. In what follows, we will describe our research setting, contextualize the thus formulated theoretical propositions, and present our hypotheses.

5.4. Dutch tax administration

This study focuses on frontline tax officials inspecting the bookkeeping records of small and medium sized enterprises. Under the heading of the so-called ‘horizontal supervision’ approach, the Dutch tax administration has embraced responsiveness and trust towards entrepreneurs as essential ingredient for compliance (Gribnau, 2007). This horizontal policy encourages officials to assess tax returns on their acceptability, rather than their mere correctness. This means that officials are encouraged to collect ‘sufficient’ information to make a judgment, and have ‘to do enough work, but not too much’ (Belastingdienst, 2016, p. 4). This practically means that officials are encouraged not to start their inspection with the assumption that it will probably be wrong, not to do their utmost to find even the smallest flaws, and not to enforce the maximum financial correction when it has been found that entrepreneurs just made a mistake and express their goodwill to change. As a consequence, assessments of entrepreneurs’ intentions and competences are part and parcel of tax officials’ judgments. The standards to assess tax returns, thus, have become less predetermined, and more dependent on officials’ assessments.

Tax officials’ evaluations of entrepreneurs’ trustworthiness is central to this study. In determining the acceptability of entrepreneurs’ tax returns, tax officials look at what is presented to them in terms of bookkeeping records, but also at whether entrepreneurs are trustworthy, in order to make inferences about the credibility of what is presented. They generally look at two aspects of trustworthiness – intentions and competences – to look at whether some sort of fraud might be involved, or whether it is a mere fault. This, in turn, influences officials’ willingness to reach a compromise, and the height of the possible fine. Within this study, we aim to cover both the evaluation of the trust that can be vested in the entrepreneur, as the evaluation of the enterprise as a whole, since these are the core evaluations tax officials make in their daily work. We are furthermore interested in tax officials’ intention to more critically scrutinize the case at hand, since this is the main decision determining whether officials will intensify their inspection or not.

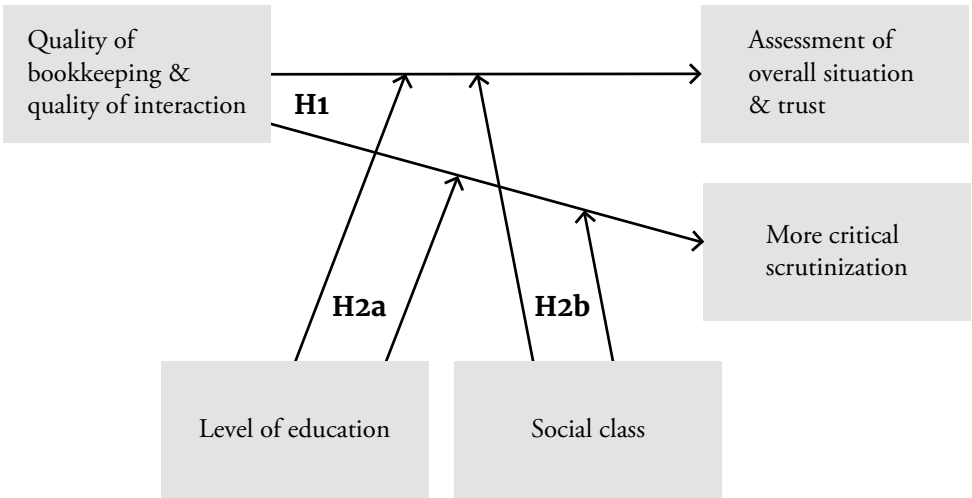
In this study, the focus is on the effect of status characteristics on the interpretation of signals indicating low or high quality of the bookkeeping and interaction. The status characteristics focused on are entrepreneurs’ social class and level of education. The study reported in the previous chapter has suggested these attributes play a role in frontline tax officials’ evaluations. Whereas these characteristics tend to overlap, they are often mentioned separately by tax officials. The findings indicated that these characteristics carry along specific expectations regarding entrepreneurs’ intentions and competences. These characteristics are moreover associated with more generic cultural beliefs that are shared by society at large. Lower educated people are viewed as generally less competent than higher educated people; a distinction which informs hiring decisions at the labor market. Although level of education is generally perceived as legitimate ground to distinguish job applicants, its relevance to street-level law enforcement is less obvious. Furthermore, lower social classes are generally perceived as less competent and in need of help (e.g. Harrits & Møller, 2014; Dubois, 2010). These status characteristics are likely to influence officials when they need to assess intentions and competences.

Based on the findings of the previous chapter, this study distinguishes two sources of attributes which serve as independent variables. Besides looking at characteristics of the bookkeeping records, and in particular how this is presented, tax officials also take into account entrepreneurs’ demeanor in the interaction to assess whether the tax return is acceptable, i.e. whether what is presented and found is credible. For this reason we distinguish quality of the bookkeeping and quality of the interaction as determinants of officials’ judgments.

5.5. Hypotheses

Within this section we formulate two sets of hypotheses. The first concerns the influence of the quality of the bookkeeping and the quality of the interaction on officials’ evaluations. Secondly, we will discuss our hypotheses on the moderating effect of entrepreneurs’ status characteristics on officials’ evaluation of quality of bookkeeping and interaction signals. Figure 5.1 portrays the conceptual model and the corresponding hypotheses.

Figure 5.1. Conceptual model with hypotheses



Since Lipsky’s seminal work on street-level bureaucrats (1980), scholars have been studying the factors influencing street-level decision making. In their study on cops, teachers and counselors Maynard-Moody and Musheno (2003) show that these street-level workers are not only state-agents implementing rules and procedures, but also citizen-agents, who take the interaction with citizens into account in their decision making. The street-level bureaucracy literature and the literature on regulatory encounters provides evidence that street-level officials not only look at characteristics related to their core task, but also at how citizens behave in the interaction, to make judgments (Maynard-Moody & Musheno, 2003). The latter authors show how street-level bureaucrats rather respond to co-operative citizen-clients than to manipulative and over-demanding citizens. Nielsen (2007) shows that the higher the level of communication (in frequency and quality), the more lenient an inspector is. Therefore, we expect that the higher both the quality of the bookkeeping and the interaction, the more positive officials’ evaluations will be.

H1: Cues indicating a good quality of bookkeeping and a good quality of interaction will have a more positive effect on officials’ evaluation of trustworthiness and overall situation, and will have a more negative effect on officials’ inclination to more critically scrutinize the entrepreneur, than cues indicating a bad quality.

Secondly, we formulate hypotheses for the indirect mechanism, which is this study’s particular contribution to the public administration literature on stereotyping at the frontline. Based on

our previous exploratory study (chapter 4) we expect that frontline tax officials use double standards to evaluate entrepreneurs. That study has suggested that differential evaluations are based on cultural beliefs about professions involving either manual or mental labor, about different levels of education and different 'classes' in society. An example mentioned within this study is the differential evaluation of 'wrongly declared turnover tax': a 'high-level' mayor, it is held, is to blame, since he should have known, whereas a shoemaker, is not to blame, because he is just incompetent. Another respondent distinguishes status groups according to their alleged intentions, and argues that residents of mobile homes cannot and do not want to keep their records properly, whereas manual workers simply do not have the skills. Another example is that of the lower educated or the 'underclass of society' who are assigned bad intentions in case of wrongly kept records, whereas the intentions of a higher educated entrepreneur in a similar situation are described as good intentions that have gone bad (*ibid.*).

These findings thus suggest that double standards are used, but they are less straightforward about the directions in which these work. In some instances, the higher status entrepreneur is evaluated stricter, whereas in other instances the lower status entrepreneur is evaluated stricter. This could be due to the qualitative and exploratory nature of that study, which did not allow us to single out and keep constant certain variables. In this current study the independent variables will be manipulated, allowing us better to assess the validity of double standards theory. In line with double standards theory and the findings of our previous study, we expect that entrepreneurs' level of education and social class serve as moderating contexts, influencing the strength and possibly also the direction of the effects of signals on the officials' evaluations. The previous study (chapter 4) has shown that a lower level of education is often associated with diminished expectations about entrepreneurs' competence. Therefore, we expect that the same situation is evaluated stricter (i.e. the evaluations will be more negative) for lower educated entrepreneurs than for higher educated entrepreneurs:

H2a: Cues of both quality of bookkeeping and quality of interaction will be evaluated stricter for the lower educated entrepreneurs than for the higher educated entrepreneurs.

Moreover, tax officials sometimes associate entrepreneurs from a lower social class not only with lower levels of competence, but also with bad intentions; i.e. entrepreneurs who try to withhold tax money (see findings chapter 4). Bookkeeping records that seem acceptable at first sight, then, could also be feigned. It is likely that such suspicions about social class influence the interpretation of signals, such as bookkeeping and interaction characteristics. For this reason, we expect a moderating impact of social class on the effect of quality of bookkeeping and quality of interaction cues as follows:

H2b: Cues of both quality of bookkeeping and quality of interaction will be evaluated stricter for entrepreneurs from a lower social class than for entrepreneurs from a higher social class.

Based on hypotheses 2a and 2b we thus expect that similar scenarios will be more negatively evaluated for entrepreneurs with a lower level of education and from a lower social class than entrepreneurs with a higher level of education and from a higher social class.

5.6. The policy-capturing study

To examine whether officials evaluate similar evidence differently for entrepreneurs from different status groups, this study conducted a policy-capturing study. The policy-capturing design allows for studying how decision makers use information in evaluative judgments (Aiman-Smith et al., 2002; Karren & Barringer, 2002). It involves letting respondents judge a relatively large set of hypothetical, but realistic scenarios in a row, with each scenario being composed of a distinct combination of cue values. Subsequently, respondents' evaluations are regressed on the cue values, which enables researchers to assess the relative weight of the various cues in evaluations.

We chose for a policy-capturing design because it allows for the study of stereotyping by officials in a context that resembles real-life decision making. Policy-capturing studies are typically more realistic than laboratory experiments where respondents are removed from their natural environments and typically evaluate only one scenario (Aguinis & Bradley, 2014). Whereas classical experiments measure officials' first stereotypical reactions, the question remains whether these studies actually capture officials' judgments in work situations or rather first impressions they share with other people in general. The policy-capturing method has better external validity, because it allows respondents to adjust their evaluations to prior evaluations. Evaluating various cases is what tax officials do on a weekly basis. Decisions about these cases are not made in a vacuum, but compared to each other. Policy-capturing studies thus resemble officials' actual work situations better, since such designs allow for assessments of multiple scenarios and comparisons between scenarios. Since respondents are asked to make judgments about scenarios including multiple cues, the policy-capturing study reduces, to some extent, the possibility for respondents to give strategic answers (Karren & Barringer, 2002).

The policy-capturing design furthermore allowed us to study different combinations of stimuli and multiple decisions, whereas traditional experimental designs can only study a limited amount of decisions. Moreover, the policy-capturing design provides a relatively high degree of control over confounding factors, because of its full factorial design. Because respondents in our study evaluated all possible combinations of the different cue values, the independent effects of each value could be assessed. Within traditional experiments, there typically is more uncertainty regarding possible other explanatory factors (Aiman-Smith et al., 2002).

Design and scenario construction

Each scenario entailed a value of the four cues (quality of bookkeeping, quality of interaction, level of education and social class). This study employed a full factorial design, which resulted in a total amount of 36 scenarios ($2 \times 2 \times 3 \times 3$). Each respondent was asked to evaluate 40 scenarios, including four duplicated scenarios. Whereas reliability is a necessary condition for the validity of measures, Karren and Barringer (2002) noted that few published policy-capturing studies analyzed the reliability of evaluators' judgments. The authors recommend that replicating four scenarios may serve as feasible test-retest check of the judgments. Our 9:1 scenario-to-cue ratio meets the minimum ratio of 5:1 as suggested by Cooksey (1996). The scenarios were presented in narrative form. In order not to exhaust our respondents, we constructed the scenarios in such a way as to only include the necessary information needed to make a judgment. We undertook 10 test interviews to improve our scenarios and operationalization of cues, aiming for an optimal balance between realism and feasibility. Appendix 3 presents an example of a scenario used.

Cue development and operationalization

For each cue we developed several behavioral statements that represented different levels of the respective cue. The choice for these values is based on our previous in-depth study on signals of entrepreneurs' trustworthiness and untrustworthiness (chapter 4), and also on 10 test interviews with tax officials. During these interviews it was assessed how statements were interpreted, which were refined or adjusted with the help of respondents' input. With regards to entrepreneurs' level of education, we chose to explicitly state the level of education (either low or high) as an impression acquired during the audit, since that is typically the way officials express their sense of an entrepreneurs' cognitive abilities.

The concept of social class is broader than socioeconomic class, since it not only refers to people's economic position in society, but also more broadly to sociocultural aspects such as lifestyle and behaviors (e.g. Harrits & Møller, 2014). In this study, we distinguished between a low and a middle-high social class. At first we tested a cue distinguishing between two known areas within the respective cities where the enterprise allegedly was located, of which one was known for its socioeconomic problems and the other was in the wealthier city center. However, since the areas were not known to all respondents, we had to develop other indicators. Therefore we chose to present pictures of streets where the enterprises allegedly were located. The pictures indicating a lower social class show multicultural streets with dilapidated buildings and poorly kept streets, whereas the pictures indicating a higher social class show well-kept streets, with well-maintained buildings and mainly white pedestrians.

For both quality of bookkeeping and quality of interaction we developed three levels, ranging from low to high quality. For the statistical analyses, the cues 'quality of bookkeeping' and 'quality of interaction' were dummy coded. The lowest levels of these cues were used as reference categories. For the three dependent variables – assessment of trustworthiness, overall judgment of the situation and intention to more critically scrutinize – we developed three items. See appendix 4 for the operationalization of all our variables. Table 5.1 shows the descriptive statistics for the dependent variables. The correlations of the independent and dependent variables can be found in appendix 5. Although the dependent variables are highly correlated, the subsequent analyses are performed for each dependent variable separately, because they measure different judgments; the variable 'appears okay' captures a general impression of the situation, the variable 'trust' measures an interpersonal judgment and 'more critical scrutinization' measures a behavioral intention.

Table 5.1. Descriptive statistics dependent variables

	N	Mean	SD
Evaluation 'appears okay'	828	3,11	1,428
Evaluation 'trust'	828	3,26	1,297
Evaluation 'more critical scrutinization'	828	5,13	1,279

Respondent selection and data collection procedure

In line with the aim of this study, we selected respondents who work with the 'horizontal supervision' policy and have face-to-face contact with entrepreneurs as part of their work. Managers of two different tax offices in two cities in the south of the Netherlands were approached, and

both were willing to cooperate with us by requesting their employees to participate in our research. 36 respondents were willing to participate. With 10 of those we conducted a test interview and with 26 we conducted the final study. For all the statistical analyses we only included respondents who had reliable response patterns, i.e. a correlation between the replicated and original scenarios of above .50 ($p < 0.10$). This resulted in a dataset with 23 respondents and 828 evaluated scenarios in total. Each row in our dataset represented an evaluated scenario. Five of the 23 respondents are female, 18 are male. Only one respondent was born in a non-western country. With regards to tenure at the time of data collection, four respondents had been in service for 3 years or less, eight respondents had been employed by the tax administration between 10 and 30 years, and 11 respondents had been in service for over 30 years.

Because the evaluation task requires respondents to invest time and effort, we decided to conduct the study within an one-on-one interview setting. In doing this, we could invest in the relationship with respondents, and enhance their motivation to participate. The first author conducted all the interviews, and the same procedure was followed within each interview. Small breaks were introduced at fixed times, to prevent respondents from getting exhausted. Appendix 6 provides a description of the procedure. After the final part of the evaluation task, respondents had the opportunity to reflect upon their experiences. This also offered us the opportunity to assess how respondents interpreted certain indicators and questions. These interviews made clear that the photos indicating low and middle-high social class were interpreted as intended.

5.7. Findings

In what follows, we will first describe the patterns of scenario evaluation found at the individual level. Second, we will test our hypotheses by multi-level analyses. Thirdly, we will use our reflection interview data to interpret the findings that were inconsistent with our hypotheses.

Individual-level exploration

In order to explore the scenario evaluations, we first conducted quantitative analyses at the individual level. IBM SPSS (version 24) was used for the analyses. We explored the direct and interaction effects on the evaluations for each respondent separately, by conducting analyses of variance. Differences across respondents were found in the patterns of direct and interaction effects involving the two status characteristics. For only five of the 23 respondents, entrepreneurs' level of education had a significant direct effect on one or several of the evaluations. No significant relations were found between social class and respondents' evaluations.

For five respondents, significant moderation effects were found. These interactions all involve a moderating effect of level of education on the relationship between a value of either quality of bookkeeping or quality of interaction with one of the evaluations. The directions of these interaction effects differed across respondents. This means that, depending on the respondent, cues of quality of bookkeeping and quality of interaction were evaluated either more negatively or more positively for the lower educated entrepreneur. We can conclude from this first exploration that for the majority of respondents no direct and interaction effects of status characteristics seemed to be at play. However, since the same analysis was repeated 23 times, the five significant interaction effects found could also have occurred by chance. Because the evaluated scenarios are nested within respondents (and observations are thus not independent), multilevel analyses were required. We estimated a maximum likelihood random intercept, fixed slopes model. We allowed respondents to vary on the dependent variables 'trust', 'appears okay' and 'more critical

scrutinization' at baseline from one another. In this model, the slopes were fixed, since we are interested in the effects of the cues (level-one units) and their interactions and not in whether these effects differ among respondents (our level-two units). Since our explanatory variables are not defined at level two, and statistical inference is only directed at respondents in our sample, a fixed effects model is appropriate (Snijder & Berkhof, 2007). Moreover, fixed effects estimates 'achieve a better control for unexplained differences between level-two units' (ibid., p. 143).

Multilevel analyses

Table 5.2 presents the results of the multilevel analyses of the direct effects of the cues and the interaction effects involving the two status characteristics on all three dependent variables. For each dependent variable, we also tested an empty model to model the random effect of respondent. For the dependent variable 'appears okay' the intraclass correlation was 0.1211 ($0.251/(0.251+1.786)$) which indicates that around 12% of the variation in the evaluation is accounted for by the respondents. For 'trust' this correlation was 0.1892 ($0.318/(0.318+1.363)$) which indicates that around 19% of the variation is accounted for by respondents. The intraclass correlation for 'more critical scrutinization' was 0.1053 ($0.172/(0.172+1.462)$); around 11% of the variation is explained by respondents. For all three dependent variables, the significant estimates of variance indicate that the intercepts vary significantly across respondents. Hence, a multilevel analysis is warranted.

Model 1 added the four cues. In line with hypothesis 1, both 'missing invoices' and 'invoices in order' have a positive effect on the evaluation of the overall situation when compared to 'barely any records'. For 'more critical scrutinization' these effects are negative and also statistically significant; the results indicate that the worse the quality of the bookkeeping, the more respondents are inclined to more critically scrutinize the entrepreneur. Regarding the quality of interaction, 'to the point' has a positive effect on 'appears okay' and 'trust' when compared to 'avoids contact'. Contrary to our expectation, 'dodging around question' has a negative effect on 'appears okay' and 'trust' when compared to 'avoids contact', but this effect is not significant. Again, for 'more critical scrutinization' these effects are reversed. This means that respondents are less inclined to more critically scrutinize the entrepreneur when s/he gives to the point answers, than when s/he avoids contact. 'Dodging around question' has a slightly more positive effect than 'avoids contact', but this effect is not significant. There were no significant direct effects of level of education and social class on each of the evaluations. For 'appears okay', adding the four cues accounts for 55.4% of the within respondent variability, and resulted in a significantly better fit of the model; the deviance decreased with 649,248 ($df=6$; $p<0.001$). For 'trust', 42.8% of the within respondent variability is explained by the cues. The deviance decreased significantly with 450,153 ($df=6$, $p<0.001$). Adding the four variables accounts for 50.4% of the within respondent variability in 'more critical scrutinization'. The deviance decreased significantly with 564,214 ($df=6$, $p<0.001$).

Model 2 added the interaction effects in order to test whether values of quality of bookkeeping or quality of interaction were evaluated differently for different status group entrepreneurs. Overall, one significant interaction effect was found for 'appears okay'; 'dodging around question' seems to be differently evaluated for lower educated entrepreneurs than for higher educated entrepreneurs. For 'trust' and 'more critical scrutinization' no significant interaction effects were found. Contrary to our hypotheses, no significant interaction effects were found for social class. For none of the dependent variables, model 2 led to a significantly better fit of the model. In order to check whether adding the significant interaction effect alone would increase the fit of the model for 'appears okay', we checked whether a new model with only the direct effects

Table 5.2. Multilevel analyses of direct and interaction effects

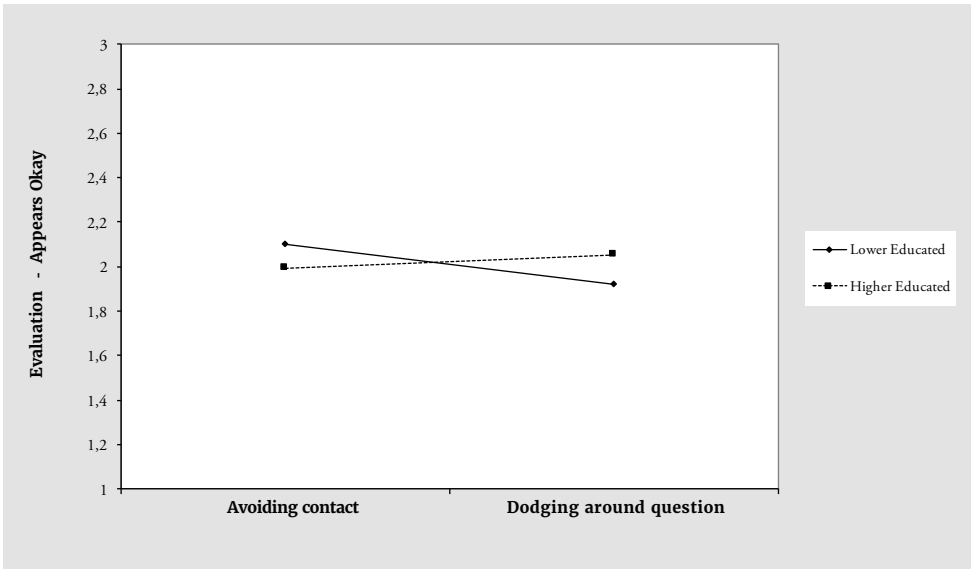
	DV: Appears okay		DV: Trust		DV: Critical scrutinization	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Intercept	2.030*** (0.137)	2.076*** (0.163)	2.450*** (0.145)	2.494*** (0.169)	5.978*** (0.120)	5.940*** (0.146)
Quality of bookkeeping						
Barely any records	Ref	Ref	Ref	Ref	Ref	Ref
Invoices missing	0.406*** (0.076)	0.428*** (0.131)	0.330*** (0.075)	0.337** (0.130)	-0.283*** (0.072)	-0.348** (0.125)
Invoices in order	2.087*** (0.076)	2.072*** (0.131)	1.406*** (0.075)	1.312*** (0.130)	-1.801*** (0.072)	-0.176*** (0.125)
Quality of interaction						
Avoids contact	Ref	Ref	Ref	Ref	Ref	Ref
Dodges around question	-0.054 (0.076)	-0.192 (0.131)	-0.083 (0.075)	-0.141 (0.130)	0.083 (0.072)	0.199 (0.125)
To the point	0.775*** (0.076)	0.768*** (0.131)	0.917*** (0.075)	0.931*** (0.130)	-0.591*** (0.072)	-0.562*** (0.125)
Level of education						
Low	Ref	Ref	Ref	Ref	Ref	Ref
High	0.053 (0.062)	-0.075 (0.138)	-0.087 (0.061)	-0.138 (0.137)	-0.017 (0.059)	0.053 (0.132)
Social class						
Low	Ref	Ref	Ref	Ref	Ref	Ref
High	-0.043 (0.062)	-0.007 (0.138)	-0.014 (0.061)	-0.050 (0.137)	0.051 (0.059)	0.058 (0.132)
Two-way interactions						
Invoices missing* level of education	-	0.014 (0.152)	-	-0.036 (0.150)	-	0.087 (0.145)
Invoices in order* level of education	-	0.130 (0.152)	-	0.145 (0.150)	-	-0.094 (0.145)
Dodges around question* level of education	-	0.254 † (0.152)	-	0.138 (0.150)	-	-0.210 (0.145)
To the point* level of education	-	0.014 (0.152)	-	-0.094 (0.150)	-	0.007 (0.145)
Invoices missing* social class	-	-0.029 (0.152)	-	0.022 (0.150)	-	0.043 (0.145)
Invoices in order* social class	-	-0.101 (0.152)	-	0.043 (0.150)	-	0.022 (0.145)
Dodges around question* social class	-	0.022 (0.152)	-	-0.022 (0.150)	-	-0.021 (0.145)
To the point* social class	-	0.000 (0.152)	-	0.065 (0.150)	-	-0.065 (0.145)
-2 Log Likelihood	2222,041	2216,909	2207,449	2202,954	2137,851	2133,073
Df	9	17	9	17	9	17
X2-change in comparison to previous model	-649,248***	-5,132	-450,153***	-5,222	-564,214***	-4,778
Variance within respondents	0.797***	0.792***	0.779***	0.775***	0.725***	0.721***
% explained variance	55.4%	55.7%	42.8%	43.1%	50.4%	50.7%
Variance between respondents	0.279**	0.279**	0.334***	0.334***	0.193**	0.193**
% explained variance	25.0%	26.1%	30.0%	30.1%	21.0%	21.1%
N (scenarios)	828	828	828	828	828	828
N (respondents)	23	23	23	23	23	23

Note: standard errors in parentheses. Significance levels: † $p \leq 0.10$; ** $p \leq 0.01$; *** $p \leq 0.001$

and the significant interaction effect would significantly decrease variance. In this new model -2 Log Likelihood was 2218,53, and X2-change was -3,511 compared to model 1. This model significantly improved the fit ($df=1$, $p<0.10$).

Figure 5.2 plots the significant interaction effect and shows that, in line with our hypothesis, a lower educated entrepreneur is judged slightly more negatively when dodging around a question than a higher educated entrepreneur. When an entrepreneur is avoiding contact, this is evaluated slightly more positive when s/he is a lower educated entrepreneur, than when s/he is higher educated. Whereas there is no significant direct effect of level of education, there is a significant, moderating effect of level of education. The difference is small relative to the scale on which the dependent variable is measured (smaller than .2 on a 7-point scale). However, the difference is larger when compared to the variance of 2.039 of ‘appears okay’, indicating a tight distribution of scores.

Figure 5.2. Graph of interaction effect Dodging around question * Level of education on ‘appears okay’



Interview data

The subsequent interview allowed us to gain insight in how respondents experienced evaluating the scenarios, and how the cues and questions were interpreted. Generally, respondents experienced no difficulty in evaluating the scenarios. Some respondents noted that the scenarios looked like each other, and that reality is more complex. In reality, for instance, they also look at what people say and not only at how the interaction unfolds. Yet, the presented cues gave them sufficient grounds to make evaluations. Also, some respondents mentioned their response pattern became less extreme throughout the evaluation task.

We moreover relied on the interview data to provide possible explanations for the findings that were inconsistent with our hypotheses. Contrary to our hypothesis, we found that when an entrepreneur is avoiding contact, this is evaluated slightly more positive when s/he is a lower educated entrepreneur, than when s/he is higher educated. A statement by one of our respondents could offer an explanation for this. He argued that when a lower educated person does not

seek contact this could have to do with insecurity, whereas a higher educated person has better interpersonal skills and is less insecure. As a consequence, the official starts to 'get suspicious' when a higher educated entrepreneur avoids contact. In this case, a higher expectation leads to a stricter evaluation when evidence for low competence is encountered than in case of low expectations. This could be a possible explanation for our 'reversed double standards' finding.

Moreover, some respondents mentioned they deliberately tried not to look at the photos and/or entrepreneurs' level of education. One respondent for instance argues that the photos may lead to expectations, and 'you look at it, but you try to block it'. Another respondent argued he learned to suppress his first impressions, in order to be as neutral as possible. Again other respondents argued that one needs to be careful with presumptions, since they do not have to be true. Some say these aspects are not supposed to play a role and are not really relevant, but that they sometimes do give a first impression. One respondent mentioned he tries to be aware of his own prejudices, and always tries to postpone first impressions, but that he does not want to be naïve either. Although trying to be nonbiased, most respondents at the same time associated specific expectations to either lower or higher status groups. E.g. 'I expect more from the higher educated, and less from the lower educated', or 'higher educated rather have a negative impact; they are more able to cheat than the lower educated'. This indicates that although some respondents learned to block their prejudices or postpone their first impressions, they can involuntary play a role. Respondents who argued they tried to not to let themselves be influenced by presuppositions or prejudices, likely also try to do this in their actual work. This may be an explanation for the nonsignificant interaction effects.

5.8. Conclusion & discussion

This study examined whether officials use double standards in evaluating entrepreneurs during inspections. It provided a first step in testing the explanatory potential of the sociological double standards theory in a public administration context. Using a policy-capturing design, this study tested whether situations involving entrepreneurs with a lower level of education and from a lower social class are evaluated more negatively than similar situations involving entrepreneurs with a higher level of education and from a higher social class. Our hypotheses were partly confirmed. Most values of quality of interaction and quality of bookkeeping, except for dodging around the question, had a significant effect on the evaluations. With regards to the double standards propositions, we found that when a lower educated entrepreneur dodges around a question this is evaluated slightly more negatively than when a higher educated entrepreneur dodges around a question. We also found evidence for the reversed practice: when a higher educated entrepreneur avoids contact this is evaluated slightly more negatively than when a lower educated entrepreneur avoids contact. This finding underlines the importance of studying indirect mechanisms of stereotyping, especially since we did not find any direct effect of status characteristics on the evaluations.

Whereas our prior qualitative study suggested tax officials may use double standards, most of the interaction effects in this study were nonsignificant. This could have different reasons. First, due to a lack of control in the qualitative study, the situations mentioned could have been less comparable. Interviews are moreover more suitable to study stereotypical reasoning than indirect ways of stereotyping. Second, within the qualitative study the use of double standards was only found for a minority of respondents. This does not necessarily translate into statistically significant interaction effects. When compared to the direct effects of most of the cues, the significant interaction effect is moreover only small in size. This is not surprising since quality

of bookkeeping and quality of interaction are deemed essential for evaluating the acceptability of tax returns, while entrepreneurs' level of education is not. More interestingly, whereas we did not find any direct effect of level of education on the evaluations, we did find it could affect frontline evaluation in combination with other signals. These differences can have a large impact on the further involvement of an inspection and decisions being made. It could make a difference between giving someone the benefit of the doubt or not. This frontline practice may harm equal treatment, and have lasting consequences for citizen-clients.

Our findings have several theoretical implications. First, they show that stereotyping by frontline officials could work more indirectly than is hitherto assumed within the street-level bureaucracy literature. Studies have shown that street-level bureaucrats rely on stereotypes in decision making as a way of coping with time pressures and high workloads (Lipsky, 1980; Andersen & Guul, 2016). These studies suggest citizen-clients' belonging to social groups serve as shortcuts to their supposed identities. Our study indicates that frontline officials employ an indirect way of stereotyping in which citizen-clients' belonging to a social group serves as frame that influences the interpretation of other signals. In fact, our multilevel analyses have shown that entrepreneurs' level of education does not have a direct effect on the evaluations, but has an effect on one of the evaluations in combination with another signal. This subtler way of stereotyping calls for research approaches that take into account how officials interpret clusters of signals.

Our study has furthermore found evidence for the use of double standards in different directions. Findings point out that the standards can be stricter for the low status entrepreneur and more lenient for the high status entrepreneur, or the other way around. In this study, 'avoiding contact' was evaluated stricter for higher educated entrepreneurs, whereas 'dodging around question' was evaluated stricter for lower educated entrepreneurs. In line with our double standards proposition, not giving answers to questions may be interpreted stricter for lower educated entrepreneurs because it is consistent with the lower expectations officials have of their competences. A possible explanation for the finding that works in the opposite direction could be that inferences about different properties are made for the different status groups. Our qualitative data suggests that a lower educated entrepreneur who avoids contact is associated with mere incompetence in communicating, whereas this is seen as a signal for bad intentions for higher educated entrepreneurs, who are expected to have these communication skills. Foschi (2000) refers to the latter as 'reversed double standards', which has been advocated by some as a means to change the status quo. Although this might be experienced and proposed by officials as more fair, it reinforces the assumption that lower status citizen-clients cannot meet the universalistic standards and therefore have to be treated more leniently (*ibid.*). Either way – in receiving a stricter or more lenient treatment – lower status groups are treated as inferior.

Following up on our findings, future research should examine how organizational socialization of public officials affects their use of double standards. Especially since some respondents suggested they have learned to block prejudices or postpone their first impressions, there are indications that organizational socialization may work to neutralize the effects of stereotypical expectations and concomitant double standards. In fact, taking into account the influence of organizational socialization, but also other background characteristics of public officials, on the use of double standards, would contribute to the development of a theory aimed at explaining the extent to which double standards are used.

Our findings also have implications for new models of governance that have come to embrace street-level officials' own professional judgments as essential for decision making. Within models promoting trust between officials and citizens, officials have to work with rules and legisla-

tion that grant them more discretion to rely on their own interpretations in decision making. Within our case, the question has shifted from ‘is it correct?’ to ‘is it acceptable?’, thereby allowing officials to look at entrepreneurs’ demeanor and at whether they appear trustworthy. Our study has shown that, in such a context, officials sometimes use double standards in evaluating citizen-clients. Whereas these new governance models allow frontline officials to be more responsive and – in our case – to get citizen-clients more compliant, this way of working also has implications for consistent and equal decision making (see also Piore, 2011; Rutz et al., 2015).

This study’s approach to examining stereotyping moreover has different advantages but also drawbacks when compared to experimental research designs using control and treatment groups. Recent experimental studies have found evidence for direct effects of stereotypes, such as ethnicity, on decision making (e.g. Andersen & Guul, 2016). We did not find such direct effects. Rather than making statements about which findings are more true, it is more fruitful to reflect on the implications of using different methods. Whereas the classical experiments do primarily measure officials’ first stereotypical reactions, the question remains whether these studies actually capture officials’ judgments in work situations, or their first impressions as human beings. Policy-capturing studies probably resemble officials’ actual work situations better, since such designs allow for assessments of multiple scenarios and comparisons between scenarios. As such, respondents have more opportunity to reflect on their first impressions and adjust their responses accordingly. However, this seems to accord with officials’ daily practice in which they try not to rely on their prejudices. An interesting venue for future research would be to analyze whether and how officials try to make their decisions consistent with prior decisions, by specifically looking at carry-over effects.

This study also has some limitations. First, this study does not allow for generalization to a larger population. We only had a small sample that was not selected on grounds of representativeness for a larger population. Yet, our main aim was to theoretically generalize: we tested the validity of the double standards theory in a new context, that is, street-level decision making. It is highly likely that our main finding that in some occasions officials use double standards is generalizable to comparable frontline domains where rules and guidelines have become less clear-cut and there is more room for officials’ interpretation. Future research should assess the prevalence of this practice among larger populations and different types of street-level officials. Second, because we had many conditions and only a small sample, we could not control for possible order effects. Therefore we kept the scenario order constant for each respondent. By using larger samples and less conditions future research could disentangle cue effects from possible order effects by randomizing the order of scenarios.

Third, the way cues were operationalized could have impacted our findings. Level of education as a signal for competence, for example, was given as an impression acquired through the inspection, and not measured by more implicit indicators. This could have raised respondents’ awareness about the focus of our study. Using more fine-grained indicators for level of education could have resulted in better identifying interaction effects. Our cue of social class, as a signal for intentions, furthermore, portrayed not only indicators of wealth and maintenance of streets, but also of ethnicity. While these often tend to go together, they are not the same. Our cue thereby grasped a broader stereotype around social class. Future research could disentangle these indicators and measure the effects of social class and ethnicity separately.

Fourth, because respondents were asked to evaluate a fairly large amount of scenarios, respondents learned about their own response patterns and the manipulated cues, and could have adjusted their responses accordingly. Although this learning effect may indeed have occurred, this probably resembles tax officials’ daily practice where they have to inspect multiple cases

on a monthly and sometimes weekly basis, and compare cases to make consistent decisions. Hence, within an experimental research design where respondents only evaluate one scenario, it is likely that there would be more and stronger evidence for the use of double standards. Yet, findings of such experiments are less generalizable to real-life settings, where officials attempt to make consistent and fair decisions. Moreover, since our study still found evidence for the use of double standards, it is likely that the amount of information in vignettes made it difficult to give strategic answers and that the trust established in the one-on-one setting made respondents feel comfortable in making honest evaluations. Future studies on frontline stereotyping could compare different methods, such as policy-capturing and experiments with treatment and control groups, to study similar research questions. In doing this, the specific contributions of each method to the study of stereotyping could be assessed and compared.

This study has shown the added value of using a policy-capturing design to examine officials' implicit use of stereotypes in decision making without stripping it of the broader decision making context. However, while the study resembles real-life settings, the scenarios are still hypothetical and compromise the complexity of real-life frontline decision making. Scholars interested in studying indirect stereotyping could consider conducting field experiments, which typically have better external validity. However, such studies are more difficult to conduct. Either way, this study has suggested that citizen-clients' status characteristics may affect the standards officials use to interpret information, without necessarily affecting their evaluations directly. This finding calls for research approaches and methods that are able to grasp this indirect, but pervasive, form of stereotyping.

Chapter 6

Conclusions and discussion

6.1. A study on bureaucratic uncertainty

Uncertainty reduction is central to the ideal typical form of bureaucracy (Gajduschek, 2003; Groeneveld, 2016). Obedience to strict rules, a clear hierarchy, and formalistic impersonality are typical characteristics of bureaucratic organizations (Weber, 1978). With regards to public organizations, these characteristics guarantee that higher order commands are fulfilled, and hence, the general interest is served (Gajduschek, 2003). Also, since outputs are calculable and processes controllable, uncertainty is also reduced for citizen-clients, who expect to be treated equally, and ‘sine ira et studio’, that is, without anger and fondness (Weber, 1978, p. 975). Emotions such as affection and enthusiasm are viewed as threats to bureaucratic functioning, opening the door to arbitrariness and inconsistency. The Weberian view on bureaucracy constitutes, as was also intended as such, an idealized top-down notion of bureaucracy.

Scholars have long been acknowledging that the reality of bureaucratic organizations is more complex, and that the ‘human factor’ that is involved leads to inefficiency and dysfunctions (e.g. Merton, 1940), or that bureaucratic behavior is only ‘bounded rational’ (e.g. Simon, 1976). Furthermore, scholars recognized that bureaucratic organizations needed to be flexible to some extent, in order to respond to the nonroutine cases (e.g. Evans, 2015; Lipsky, 1980). This flexibility is embodied in the role of the street-level bureaucrat, who has discretion to make decisions, within the given rules, as she or he sees fit. Scholars started studying how frontline officials use this discretion, and the factors influencing discretionary decision making (e.g. Lipsky, 1980; Prottas, 1979; Maynard-Moody & Musheno, 2003). Discretion, just like uncertainty reduction, is a top-down notion which has been of interest to a variety of research traditions within public administration, such as public management, public accountability, policy implementation and street-level bureaucracy.

The notion of discretion implies uncertainty which is deemed problematic at higher levels of the bureaucratic organization, because it entails indeterminacy of bureaucratic actions which are therefore difficult to control. At the same time, this uncertainty is sometimes framed positively, as it entails flexibility to adapt to unexpected situations, and this responsiveness, in turn, could increase effectiveness and legitimacy. Instead of taking a stance in this discussion, this dissertation has treated uncertainty as an empirical phenomenon, studied from the perspective of frontline bureaucrats themselves. That is, rather than assuming that uncertainty is part and parcel of discretion, which should be controlled, this dissertation has studied uncertainty and uncertainty reduction from a bottom-up perspective.

A bottom-up perspective to the study of frontline work is especially important since the

frontline worker and his or her interaction with citizen-clients, hence the human factor, is nowadays not only deemed necessary to implement policies, but also valued by public bureaucracies as a tool to increase responsiveness and improve service delivery (Bartels, 2014). This involves more leeway for individual bureaucrats to make judgments based on their own professional standards (Dubois, 2014; Evans, 2015). Interactions with citizen-clients are seen as sites where responsiveness can be achieved, allowing individualized judgments and decision making (Rice, 2013). In such a context with increased leeway, it is likely bureaucrats also look at particularistic citizen attributes to make judgments, and that social interactions have an impact on decision making, possibly leading to arbitrariness and inconsistency. This empirical shift towards more 'local inter-individual arrangements' (Dubois, 2014, p. 39), also requires a contextualization of the study of bureaucratic uncertainty and uncertainty reduction. This dissertation has focused on the bureaucrats working within such a context. Guided by the following general research question, this dissertation has employed a bottom-up approach to the study of bureaucratic uncertainty:

What kind of uncertainties do frontline officials – working within a bureaucratic context that has expanded frontline leeway and encourages professional judgment – experience, and how do they attempt to reduce these uncertainties?

By answering this question, this dissertation has aimed to understand uncertainty experiences from the perspective of bureaucrats themselves, and the role of social context, stereotypes and influence of double standards as particular ways to reduce uncertainty. As such, it could be assessed whether bureaucrats working with more leeway experience particular uncertainties, and how frontline behavior within such a context relates to bureaucratic notions of uncertainty such as inequality.

In order to answer the research question, this dissertation focused on a bureaucratic organization where officials have over the years been encouraged to rely on their own interpretations in making judgments. To enhance compliance, the Dutch tax authorities have adopted a policy promoting responsive law enforcement (Gribnau, 2007). Under the heading of the so-called 'horizontal supervision' approach, officials working for the Dutch tax administration have been encouraged to be responsive to individual cases. The assumption is that taxpayers are not merely being driven by economic incentives and sanctions, but by a range of other factors such as social norms, values, habits, but also by the interaction between the tax administration and taxpayer (Gribnau, 2007; Leviner, 2009; Kirchler, 2007). As a consequence, the interactional processes between tax administration and taxpayers are seen as crucial in fostering compliant behavior: 'maintaining open communication and positive and professional service, even through the toughest encounters with taxpayers, is important for the tax authority (...) because, in most cases, even when resentment, anger, and disobedience are present on the part of the taxpayer, there is also goodwill, and therefore, an opportunity to draw out the more cooperative motivational postures' (Leviner, 2009, p. 263).

This has led to several changes at the frontline. Firstly, officials auditing small and medium sized businesses are now expected to go on inspection with a trusting attitude, meaning they are encouraged not to correct every fault they find. Relatedly, officials are mandated to assess the acceptability of tax returns, rather than the mere correctness, allowing more leeway to interpret what is acceptable in practice. Thirdly, officials are allowed and even encouraged to make settlement agreements with entrepreneurs, involving negotiations with entrepreneurs.

One of the empirical chapters has focused on how social dynamics could impact decision

making about specific cases. It not only looked at interactions between officials and citizen-clients, but also at the social context of frontline teams. In order to be more responsive to complex cases, the Dutch tax administration has also set up multidisciplinary ‘take-away teams’ aimed at combating organized crime and confiscating criminal assets. It involves a collaboration between five large municipalities in the province of Brabant in the Netherlands, the public prosecutor, the police, Fiscal Information and Investigation (FIOD), the tax authorities, and the Royal Military Police. The different parties exchange information about suspects, and collaboratively make a decision. Within the Belgian Inspectorate of Social Laws, interorganizational teams have also been set up to enhance efficiency and decrease the frequency of inspections, but also to increase consistency in decision making across organizations.

Each empirical chapter addressed a different sub-question, using different methods. Using an inductive storytelling approach, chapter two studied the kinds, sources and responses to uncertainties experienced by frontline tax officials. Chapter three combined inductive interview and observation data from officials working for the Dutch tax administration and from inspectors working for the Belgian labor inspectorate, to assess the impact of social dynamics and context in different frontline constellations on decision making about citizen-clients. Chapter four studied the signals tax officials look at to categorize citizen-clients as either trustworthy or untrustworthy, and explored whether officials use differential standards in assessing similar signals of citizen-clients from different status groups. Chapter five, finally, has tested the double standards mechanism, by examining whether citizen-clients’ status characteristics affect officials’ evaluations of similar signals unequally.

The conclusions of the findings of each empirical chapter will be presented in section 6.2. After that, in section 6.3, an answer to the general research question will be provided. In section 6.4 the societal relevance of the findings will be discussed. Section 6.5 discusses the theoretical and methodological implications of the dissertation. In section 6.6 the limitations of the dissertation will be discussed, and lastly, in section 6.7, recommendations for future research will be made.

6.2. Conclusions

This section will present the conclusions of each empirical chapter separately. It will furthermore show how the research questions posed and methods used in each chapter build on the findings of the prior chapters.

Different kinds of uncertainty

Chapter two examined the kinds, sources and responses to uncertainties experienced by frontline tax officials. It analyzed thirty-seven uncertainty stories collected through in-depth interviews with seventeen tax officials. The aim was to understand uncertainty from a frontline perspective in a bureaucratic context where officials have over the years been endowed with more leeway to make judgments and decisions. The classical literature on bureaucracy and the street-level bureaucracy literature were reviewed, two notions of uncertainty were distilled – information and interpretation uncertainty – and it was concluded that a more sociological notion of uncertainty – social uncertainty – is often missing in the public administration literature. The three notions of uncertainty were used as sensitizing concepts in the analysis of the stories.

In line with the classical literature on bureaucracy (e.g. Simon, 1976) this study has found that tax officials experienced information problems. However, they were not more broadly

looking for information, but looked for proof supporting their account of what happened. The common response was looking for more proof, constructing a convincing account, and/or satisficing, that is, trying to get the account accepted with available evidence, or accepting that their account of what happened had to be adjusted. This kind of uncertainty directs attention to the role of other factors than information, such as cues and signals, in frontline officials' formation of a picture of a case.

In accordance with uncertainty as reflected upon in the street-level bureaucracy literature (e.g. Lipsky, 1980), tax officials experienced interpretation problems. These interpretation uncertainties involve difficulties with standards in making decisions. Conflicting norms and/or feelings, and vague rules and regulation yield questions as to what is the right decision. Deliberation with colleagues is a common response in such situations, since then, respondents argued, they have a back-up and have to struggle less with the situation afterwards.

This study furthermore found that tax officials face action uncertainties which involve (a fear of) losing control in interaction with the entrepreneur, or uncertainty about how to get it back. The stories showed this often happens in interaction with entrepreneurs, which can never be fully predicted. This puts officials on guard because they often need to take ad hoc action and feel their on-the-spot actions weigh heavily on the course and outcome of an interaction. In these situations officials usually do not have the time to think through their actions, or their planned actions do not suffice, requiring improvisation. This finding underlines the importance of improvisation in officials' daily work.

The study also explored the sources of these uncertainties and showed that there are structural aspects of frontline work, such as the 'unknowable truth' and the ambiguous character of rules and regulation that underpin officials' search for proof and interpretation uncertainty. In addition, stories also point to the 'horizontal supervision policy' as a source of different uncertainties. The policy has given officials more interpretation leeway, involving doubts about what is the right decision. It furthermore encourages negotiations with entrepreneurs, in which officials feel they possibly lose control over the situation. There are also external factors, such as economic hardships and the undesirability of audits on the part of entrepreneurs that underlie the latter's emotional outbursts in interactions.

Social context of decision making

Whereas the second chapter has inductively shown that officials rely on their colleagues to deal with interpretation uncertainty, the subsequent study specifically focused on exploring the impact of frontline officials' social context on frontline decision making. Using qualitative interview data of the Dutch tax administration and interview and observation data of the Belgian labor inspectorate, this study examined not only frontline officials' deliberate reliance on peers, but also the more implicit social dynamics in interactions with peers and citizen-clients. The study has shown that whereas frontline officials' reliance on peers is motivated by their wish to make less subjective judgments, implicit social dynamics among colleagues in teams and in interactions with citizens, such as social pressure or social ties, sometimes lead to decisions an individual official would or could not have made by him- or herself. The informal meetings with other professionals at 'the kitchen table' are experienced as opening up opportunities to better solve difficult cases, whereas social dynamics – such as social pressure – are experienced as problematic when the resulting decision runs counter to officials' own idea of what is the appropriate decision. Moreover, in negotiating with citizen-clients, final decisions are for a large part dependent on what is brought to the table and the involved parties' negotiation and persuasion skills.

The findings underline the importance of studying the social context of frontline decision making; social dynamics could either lead to uncertainty reduction for the individual bureaucrat, but also lead to arbitrariness in decision making, and thus, bureaucratic uncertainty. The study furthermore showed the added value of combining data from different cases, although collected independently and with different purposes, in showing the ubiquity of social dynamics in frontline decision making.

Signals and stereotypes

Using in-depth interviews and the method of storytelling, chapter four has studied how officials may evaluate similar signals differently for entrepreneurs from different social groups. In order to answer this research question, the study first addressed the question what signals officials look at to assess citizen-clients as either trustworthy or untrustworthy and how they do this. It has built on chapter two's finding which indicated that officials' image construction of citizen-clients is based on other factors than only information. Studies on service encounters have pointed to the importance of signals and cues in reducing uncertainty (Gambetta & Hamill, 2005; Mennerick, 1974). Since officials are interested in assessing citizen characteristics, in this case trustworthiness, which cannot be observed, it is held, they look at observable signals that are believed to be linked to these unobservable characteristics (Gambetta & Hamill, 2005). For this reason, this study has focused on the signals and cues officials look at to assess entrepreneurs' trustworthiness. The study has built on the findings of chapter three too, which has pointed to the subjectivism of individual officials' decision making, both in terms of differences in approaches of officials, but also in terms of the impact of relationships with citizen-clients. In such a context, it is highly likely that stereotypes also affect the standards officials use. Chapter four therefore explored the signals looked at to assess trustworthiness, and by relying on the sociological double standards theory (e.g. Foschi, 2000), assessed whether and how similar signals are differently evaluated for citizen-clients belonging to different social groups.

Addressing the exploratory question, this study found that frontline officials not only look at signals related to entrepreneurs' bookkeeping, but also their appearance, demeanor, home situations and belonging to social groups to make inferences about the latter's competences and intentions. The specific contribution to the literature on stereotyping by public officials is the finding that social typologies may also serve as frames affecting the interpretation of other signals. We have found this could work in different directions. In line with the sociological double standards theory, we have found that a signal indicating a bad quality of bookkeeping was evaluated stricter for lower status groups than a similar signal for higher status groups. One respondent, to the contrary, explicitly argues he has stricter standards in evaluating mistakes for a high status entrepreneur such as a mayor than for a lower status entrepreneur, such as a shoemaker.

The findings of this study, to conclude, show that in a context where officials' evaluation task involves considerable leeway, and is based on interpersonal assessments, officials not only look at bookkeeping-related signals, but also at particularistic attributes such as citizen-clients' demeanor and belonging to social groups. More importantly, this study suggests that this belonging to a social group not only serves as a signal (for example, 'higher educated are competent'), but could also serve as a frame influencing the interpretation of other signals (for example, 'bad filing behavior is a signal for bad intentions for the higher educated, but a signal for incompetence for the lower educated'). This finding points to the relevance for studying more nuanced (less visible) ways of stereotyping involving clusters of signals.

Double standards in decision making

In order to more systematically test the sociological double standards propositions, chapter five employed a policy-capturing study. This chapter builds on the findings of the previous chapter. Whereas chapter four explored the double standards mechanism by using interviews, and suggested double standards may indeed be used, chapter five tested this mechanism deductively, using a policy-capturing approach. The inductive approach of chapter four allowed for the study of how officials' interpretations of signals may differ across different social groups. It showed how officials may have different expectations of different social groups, involving differential evaluations of similar signals. The findings of chapter four are not conclusive about the direction in which the double standards may work; the interview data suggest that lower expectations of lower status groups may either result in a stricter or more lenient evaluation of signals of the lower status groups. Also, the use of in-depth interviews made me dependent on respondents' own stories, thereby limiting control of the research context.

The policy-capturing design of chapter five's study, to the contrary, allowed for more control of the research context. It involved letting tax officials evaluate a fairly large number of vignettes in an interview setting. The vignettes portrayed an inspection at a clothing store, and varied on cue values indicating different levels of quality of bookkeeping, quality of interaction, entrepreneurs' social class and level of education. The evaluation task involved three items measuring a global assessment of the situation, trust and intention to more critically inspect the entrepreneur. Because all possible combinations of cue values were evaluated, the independent effect of each cue on the evaluations could be assessed. The study thus examined the moderating effects of the status characteristics social class and level of education on the relationship between signals indicating quality of bookkeeping and interaction and the evaluations. To be more precise, by using multilevel analyses it tested whether cues of both quality of bookkeeping and quality of interaction were evaluated stricter (that is, more negatively) for entrepreneurs from a lower social class and with a lower level of education than for entrepreneurs from a higher social class and with a higher level of education.

To conclude, the majority of the hypotheses were not supported by our findings. Only some hypotheses were supported, indicating that some signals are, indeed, evaluated differently for lower status entrepreneurs compared to higher status entrepreneurs. One value of quality of interaction ('dodging around question') was evaluated slightly more negatively for the lower educated entrepreneur than the higher educated entrepreneur. There was also evidence for the reversed practice; a stricter evaluation of a cue for the higher educated than the lower educated. The qualitative reflection data provided possible explanations for the use of reversed double standards, but not for the stricter evaluation of lower status groups. This may indicate social desirability in the interview data. The vignette approach thus allowed for the study of the more implicit (or unconscious) use of double standards.

6.3. General conclusion

Within this section the overarching research question will be answered by combining the findings of the different empirical studies. The first part of the research question is about understanding frontline uncertainty in a bureaucratic context that has expanded frontline leeway and encourages professional judgment. The second part is about officials' particular ways of dealing with this uncertainty in such a bureaucratic context.

Understanding uncertainty

Bureaucratic uncertainty can take different forms and can be studied at different levels. Whereas literature on bureaucracy typically assumes uncertainty reduction is central to bureaucratic organizations, there has been no systematic study on the uncertainties experienced by frontline officials themselves, and how a more indeterminate bureaucratic process, by giving more leeway to frontline officials, relates to these experiences. This dissertation has found frontline officials basically experience three different kinds of uncertainty: information, interpretation and action uncertainty. Information uncertainty involves a lack of proof for one's account of what actually happened. Interpretation uncertainty involves doubt about the standards that can best be used to evaluate the situation. Action uncertainty, thirdly, involves (a fear of) losing control of the situation, which often happens in interaction with the citizen-client where immediate action is required.

The study also finds different kinds of responses: finding more evidence and writing a convincing report, deliberation with colleagues to reach 'intersubjective' account, and on-the-spot improvisation. The analysis of the sources of the uncertainties has revealed that some are inherent to frontline work, such as the indeterminate character of social interactions which are never fully predictable. Furthermore, the horizontal policy allowing more room for interpretation and negotiation also constitutes a source of uncertainty, giving officials 'room to struggle', and putting emphasis on their immediate actions in interaction with citizen-clients. As such, uncertainty experienced by frontline workers is partly linked to the flexibility built into the bureaucratic organization in order to be responsive to specific cases.

Besides uncertainty experienced by frontline officials, the dissertation has furthermore given insight in how social dynamics among officials but also in official-client interactions relate to uncertainty in outcomes of decision making. The study has shown that officials deliberately rely on their colleagues when in doubt about 'how to interpret', as to create more consistency in decision making. However, the more implicit social dynamics between officials and citizens, but also among coworkers working in teams, can lead to arbitrariness in decision making. Some officials problematize the impact of social dynamics such as social pressure and strong social ties on decision making, especially when the outcome deviates from what they themselves would decide. The study has also shown that when officials feel that strong social ties help 'to get things done' and benefit responsive decision making, this is not perceived as problematic. To the contrary, the overarching goal of the team to be responsive seems to justify, for them, their informal kitchen table conversations.

Apart from the question whether officials problematize the impact of social dynamics on decision making, it constitutes a source of uncertainty at higher levels in the bureaucratic organization. Social dynamics among peers but also between officials and entrepreneurs, sometimes led to a different treatment of similar cases, and also involved the experience of 'subjective' decision making. In such instances, social dynamics have a considerable impact on the decision-making outcome, which is thus not only determined by considerations of frontline officials. Although encouraged in some instances, as is the case with more responsive decision making in the form of interorganizational frontline teams and negotiations with citizen-clients, but also in teams set up to promote consistent decision making, bringing in social interactions as part of decision making involves bureaucratic uncertainty in the form of decision making motivated by *social dynamics*.

Dealing with uncertainty

This dissertation has also given insight in how frontline bureaucrats who have considerable leeway and are encouraged to make professional judgments, attempt to reduce uncertainty. Notions as trustworthiness and compliance play an important role in frontline decision making, at the same time, however, these categories largely remain undefined and up to the frontline official to define in practice. Moreover, such categories are not readily observable, since they are 'private' properties of citizen-clients, who consciously and/or unconsciously send signals about who they are and what they do. This study has shown that frontline tax officials look at different kinds of signals to reduce uncertainty, that is, to know more about entrepreneurs' trustworthiness. To assess entrepreneurs' intentions and competences (two dimensions of trustworthiness), the analysis of qualitative interview data has shown, officials look at bookkeeping-related signals and prior knowledge, but also at more particularistic attributes such as demeanor, characteristics of entrepreneurs' home situation, and social typologies. Some of these social typologies have a particular meaning for tax officials, such as 'the starting entrepreneurs', others are based on broader societal cultural beliefs such as 'the lower educated' and 'the foreigners'.

Findings have moreover pointed out that there are three ways in which officials come to interpret certain attributes as signals: extrapolating prior knowledge to future behavior, comparing specific cases to numeric averages (as part of accounting strategies) and to conceptions of what is normal behavior (based on own conceptions) and matching citizen cues to meaningful social typologies. Officials have indicated that the signals do not have a conclusive meaning on their own, but are only meaningful within the bigger picture, based on other signals, that has already been constructed. This indicated that signals may be interpreted differently for people from different social groups. For this reason, relying on double standards theory, this study first explored and subsequently tested whether tax officials use double standards in evaluating similar signals of citizen-clients from different status groups.

Exploring the double standards mechanism, the qualitative interview study suggested officials may use differential standards to evaluate similar signals of citizen-clients from different status groups, such as social class and level of education. In some instances, a bookkeeping-related signal was interpreted more negatively and in another instance more positively for the lower status group. Although double standards theory posits that evidence from lower status groups are evaluated more negatively, this qualitative study has also pointed to the reversed practice. Since the qualitative study could only explore the double standards mechanism inductively, a more systematic approach was needed to assess the effects of status characteristics on officials' interpretations of signals. The policy-capturing study allowed for the study of the independent effects of status characteristics by using a vignette approach. It tested the double standards propositions and suggested officials may indeed use differential standards. Although some evidence was found for double standards and reversed double standards, most hypotheses were not supported.

The findings of the dissertation thus pointed out that in a context where officials are encouraged to assess unobservable and undefined properties such as trustworthiness, officials also look at particularistic citizen attributes that are believed to be related to the unobservable properties. Moreover, officials also use their own subjective – sometimes differential – standards to interpret these signals. This dissertation has shown that in a bureaucratic context where frontline professional judgment is encouraged, uncertainty reduction has a particularistic character, in terms of differences in strategies used by officials, but also in terms of characteristics of the bookkeeping and of citizen-clients that matter in some, but not in other instances. This depends on officials' interpretations of situations. The particularistic ways of reducing uncertainty, thus, indeed al-

low responsiveness at the frontline, but are in contrast to the core characteristics of bureaucracy, such as impersonality and equality of treatment.

6.4. Societal relevance

The findings of this dissertation have a societal relevance. The study has shown that frontline workers experience different kinds of uncertainty in a policy context where they are encouraged to use their own professional judgment to assess citizen-clients. As such, this thesis has offered valuable insights to governors, policy makers and managers, who may be interested in the uncertainties that accompany certain policies and frontline work. Because of their formal accountability towards the top, they may want to reduce uncertainty with regards to policy implementation, as to ensure that political decisions are carried out. The study has pointed out that the 'horizontal policy' which encourages frontline officials' professional judgment, is a source of interpretation and action uncertainties experienced by such officials. Although negotiations with citizen-clients and emphasis on professional judgment may allow for more flexibility, and hence, tailor-made services or personalized solutions, it also entails uncertainties at the frontline that are difficult to control by frontline officials themselves, let alone by higher officials.

Moreover, the study has shown that officials in such a context use stereotypes and differential standards in evaluating citizen-clients. This often occurs implicitly. Since officials often work individually, such practices likely remain unnoticed. Although some respondents mentioned they rely on their social network when they are in doubt or feel they are 'in a tunnel vision', this social deliberation is not part and parcel of the frontline context which is studied. These findings have pointed out that reflection practices among colleagues could enhance consistent decision making, without necessarily hampering public officials' flexibility in being responsive. Moreover, the findings could also be valuable for frontline officials themselves, who – as my study has shown – often want to work without prejudices and value equality before the law. Insight in their own implicit stereotypical judgments, could be a step towards more reflective frontline judgments.

Moreover, this study has shown that social dynamics, such as strong social ties and social pressure, within frontline teams *could* hamper consistency in decision making. In such instances, specific decisions are not only made based on substantive reasons, but also motivated by social dynamics. The findings of the dissertation pointed out that respondents who work in teams where officials still have their individual tasks, social dynamics are experienced as negative, since the social pressure hampers their individual discretion. The respondent in the team with shared discretion, to the contrary, told stories about how strong social ties within the group increased the action possibilities. Although experienced as positive, the stories show how informal meetings are at odds with existing rules. This offers public managers valuable insights with regards to possible unintended consequences of using teams to either increase responsiveness or consistency in decision making.

The findings of this thesis are of general interest too. As the other side of the interaction, citizen-clients are the 'subject' of frontline decision making. Although citizen-clients are increasingly made part of decision making processes as to increase for instance collaboration, trust, or compliance, there are still aspects of decision making that remain implicit, particularistic and thereby invisible for citizens, such as uncertainties and stereotyping. This uncertainty in treatment of citizens can have a negative effect on society, but also on the functioning of democracy (Gajduschek, 2003). This study suggests there may be a tradeoff between more flexible and responsive frontline decision making and overall social utility.

6.5. Discussion

The findings of the dissertation have implications for the public administration literature and related fields. In what follows, the specific contributions to the literature and theories will be discussed, after which methodological implications will be discussed.

Theoretical implications

Firstly, the findings of this study have implications for literature on bureaucratic organizations. Within public organizations there is both a shift towards more standardization, by the use of ICT and performance requirements (Bovens & Zouridis, 2002), and towards more flexibility and responsiveness, by giving public officials more leeway to make professional judgments about particular cases (Rice, 2013; Dubois, 2014). Literature has pointed to the tension between bureaucratic consistency and flexibility, that is, between equal treatment of similar cases and bureaucratic action that is responsive to the complexity and needs of the external world (e.g. Piore, 2015; Rutz et al., 2015). Focusing on a bureaucratic organization which is fairly rationalized in terms of its procedures, but which at the same time has been encouraging public officials' professional judgment in making decisions, this dissertation has employed a bottom-up perspective to the study of bureaucratic uncertainty. This dissertation has shown that for tax officials a similar tension exists between on the one hand treating entrepreneurs equally, and making consistent decisions and on the other hand aiming to get entrepreneurs more compliant with tax regulations by not enforcing very strictly, and making settlements agreements. Although some officials aim to make judgments that are supported by colleagues, they do this on their own initiative. Furthermore, their use of differential standards mainly occurs implicitly, is therefore not experienced as problematic, and likely not responded to with a deliberate reliance on colleagues. Officials' ways of reducing uncertainty are thus largely individualized, probably resulting in disparities in frontline practices and judgment.

A bottom-up perspective to bureaucratic uncertainty has shown that the flexibility that is built into the organization to respond to particular cases and to increase compliance, also involves uncertainty at the frontline, and individualistic assessments and judgments. Rather than assuming that enabling officials to make professional judgments involves decisions made out of certainty or professional knowledge, for officials it also involves a 'freedom to struggle'. They sometimes aim for increasing consistency by relying on judgments of peers. The bottom-up perspective of this dissertation thus pointed out how officials at the frontline meaningfully deal with tensions, and find practical solutions to both make responsive and consistent decisions.

The dissertation, secondly, has made several contributions to the street-level bureaucracy literature. This branch of literature mainly focuses on describing or explaining officials' use of discretion, that is, the leeway public officials have within the given rules and regulation. Studies have focused on either the micro-level of frontline decision making, such as descriptions of public encounters between officials and citizens (e.g. Goodsell, 1981; Maynard-Moody & Musheno, 2003), the effects of client characteristics (e.g. Harrits & Møller, 2014), and the effects of official's characteristics (e.g. Oberfield, 2014), or macro-level determinants such as the influence of legislation, policies and politicians (e.g. May & Winter, 2009). What has hitherto largely remained out of sight in the street-level bureaucracy literature is the impact of the meso-level on frontline decision making, that is, the social context in which street-level bureaucrats operate. This dissertation has pointed to the role of social dynamics in decision making about citizens. Especially since social interactions between officials in frontline teams, but also between officials and citizens in the form of negotiations, have become part and parcel of frontline decision making, the study of the social character of frontline decision making

becomes an important topic of study. This study has contributed to the existing literature, by showing that officials not only purposefully rely on their colleagues in making decisions, but by pointing to the more implicit social dynamics, such as strong social ties and social pressure.

Another contribution to the street-level bureaucracy literature is this study's specific finding of action uncertainties that often accompany interactions with citizen-clients. How public officials improvise in the face of unexpected and deviant emotions and reactions of citizen-clients, is still to a great extent a black box. The literature on frontline decision making still falls within the 'implementation-control-discretion narrative' (Maynard-Moody & Musheno, 2012), focusing on how public officials apply rules and use knowledge (Wagenaar, 2004). This study has shown that when facing action uncertainties, officials often need to make on-the-spot judgments. Officials' stories bear witness to frontline action that is motivated by on-the-spot improvisation, which points to the relevance of a bottom-up approach to the study of frontline behavior and decision making. Such an approach allows for the study of more practical forms of knowledge and the values and emotions officials rely upon in responding to these action uncertainties. This is all the more important given that officials' immediate actions are often consequential for the further course of an interaction with citizen-clients and sometimes even the outcome of decision making.

These findings, furthermore, have implications for the notion of discretion at the frontline. Discretion assumes a degree of leeway and autonomy left open by rules and laws, given to public professionals who allegedly have the expertise to deal with such autonomy. However, this study has shown that this autonomy is also affected by social dynamics in different frontline constellations. Frontline decisions do not solely result from officials' individual and deliberate use of discretion, but also from social dynamics such as negotiations, social pressure and strong social ties. Moreover, the stories on on-the-spot judgments are not about applying rules and knowledge, but about 'getting control'. The discretion granted to officials is, thus, not entirely 'controlled by' individual officials, but part of the broader social context of frontline decision making. As a top-down concept, discretion does not seem to accord with the day-to-day practice of frontline decision making (Maynard-Moody & Musheno, 2012).

Thirdly, this dissertation has contributed to the research on stereotyping by frontline officials. Following Lipsky (1980) there has been much scholarly attention to the study of discrimination by street-level officials as a way of coping. The experimental research designs that are increasingly used to study street-level discrimination (e.g. Andersen & Guul, 2016) and the study of frontline behavior by relying on psychological theories (e.g. Grimmelikhuijsen et al., 2017; Tummers, 2016), typically assume and test individualist and psychological mechanisms of stereotyping and behavior. Sociological explanations of stereotyping are generally not studied. In line with sociological studies on stereotyping by public officials (Dubois, 2014; Epp et al., 2014; Harrits & Møller, 2014; Schram et al., 2009), this dissertation has studied whether and how societal stereotypes affect frontline judgment, and tested a sociological mechanism. Moreover, whereas most studies study the direct effects of citizen characteristics on decision making (e.g. Andersen & Guul, 2016), or show how officials' descriptions of citizen-clients overlap with stereotypes (e.g. Dubois, 2014), this dissertation has not only examined direct effects, but also how stereotypes may serve as moderating contexts, influencing the interpretation of other signals. By exploring and testing the sociological double standards mechanism, this dissertation has studied how stereotyping by frontline workers may be more nuanced than is hitherto assumed and studied within public administration research.

Fourthly, this dissertation has also contributed to the literature on public management. The findings of the different studies have pointed out that in a frontline context where officials

have considerable leeway to make judgments, they use stereotypes and double standards in evaluating citizen-clients. Although this study did not compare different frontline contexts, the findings suggest that the 'horizontal policy', giving officials more room for interpretation, allows for particularistic judgments. Although such policies may yield more responsive decision making, they could also hamper the fair and consistent treatment of citizen-clients. This has implications for the public management literature, interested in questions of how officials are managed to multiple ends, such as making consistent and responsive decisions (e.g. Rutz et al., 2015). Public management scholars begin to see the relevance of social context in managing frontline decision making (e.g. Piore, 2015; Rutz et al., 2015). Some officials in this study used their social network to reach an 'intersubjective' account that is backed-up by colleagues. This study has shown that officials reach out to colleagues on their own initiative and when they are in doubt. It is therefore likely that these efforts are mostly done by officials who are reflective, have a social network, and do not feel burdened to ask for help. This points to the relevance of social context, the development of frontline social networks (e.g. Hatmaker et al., 2011) and the socialization of officials (e.g. Moyson et al., 2017; Van Kleef, 2016) for the public management literature.

The accountability literature, fifthly, points to different forms in which public officials feel accountable. Accountability assumes 'a social relationship in which an actor feels an obligation to explain and justify his conduct to some significant other' (Bovens, 1998, p. 192). Hupe and Hill (2007) distinguish three forms of public accountability: public-administrative, professional, and participatory accountability. The first form involves a vertical notion of accountability, including managerial oversight, performance indicators and the law. The basis is a legitimate and authoritative jurisdiction. The second and third constitute horizontal forms of accountability. The basis of professional accountability is the 'abstract knowledge' which is part of a certain vocation, such as in medicine or police. Professionals practice accountability to colleagues both within and outside their organizations. Participatory accountability involves officials who hold clients and collaborators accountable and the other way around. This dissertation has pointed to the importance of social dynamics in frontline decision making, for which *individual* officials, strictly speaking, could not be held responsible. In situations where social pressure plays a role, for instance, officials feel they have no case-related reasons to justify the decision being made. Although the act of negotiating with clients, as an example, may be mandated by the organization (implying public-administrative accountability), the outcome is partly made dependent on the dynamics of social interactions. Therefore, it is likely that encouraging interactions and professional judgment at the frontline, does not necessarily accompany professional and participatory accountability, but could also lead to an erosion of autonomy and sense of responsibility.

The policy-capturing study, lastly, not only contributed to the street-level bureaucracy literature, but also to the double standards theory. Double standards theory has mainly been applied to study evaluations of competence and studies have generally concluded that lower status groups are commonly evaluated more strictly (e.g. Foschi, 2000; Ridgeway, 1991). This study not only studied evaluations of competence, but also intentions. The findings of the study suggest that different mechanisms for standards of competence and intentions may be at play. The interview data of the policy-capturing study suggest that officials generally have lower expectations of lower educated citizens' competences. Officials' expectations with regards to intentions, to the contrary, are generally lower for the higher educated citizen-clients. It is assumed that the higher educated citizen-clients know better how to cheat, because they allegedly have better bookkeeping know-how and skills. The higher expectations regarding higher educated entre-

preneurs' competences, the data suggest, seems to lead to lower expectations regarding their intentions, compared to the lower educated entrepreneurs.

The interview data did not yield insight in officials' evaluations of such cases, but it is conceivable that the lower expectations of lower educated citizen-clients' competence is met with stricter evaluations of, for instance, bookkeeping records that seem 'in order', compared to the evaluation of the same situation for the higher educated. However, since high expectations regarding higher educated entrepreneurs' competences seems to be related to low expectations regarding their intentions, it is also possible that evaluations of such evidence are stricter for higher educated entrepreneurs. The evaluations regarding entrepreneurs' competences and intentions are possibly lumped together in the more generic evaluations – such as trust – of this study.

The likely interconnectedness of standards to evaluate competence and standards to evaluate intentions of the higher educated may explain why this study does not find strong support for the double standards mechanism. The explanation might be that the high expectation of higher educated entrepreneurs' competence, contrary to our hypotheses, does not lead to a lenient interpretation of signals looked at in inspections, but leads to a low expectation of their *intentions*, and therefore to a strict interpretation of those signals. Therefore, differences with evaluations of signals of the lower educated may be minor or nonexistent, because these signals are, in line with the mechanism for standards of competence, strictly evaluated. The study thus points to the importance of studying the interconnectedness of different sorts of standards (for the inference of different attributes, such as competence and intentions) for evaluating evidence of different status groups.

Methodological implications

The dissertation has also made some methodological contributions to the literature. Firstly, by using an inductive research approach, relying on the methods of in-depth interviews and storytelling, this study has pointed to the importance of frontline officials' experiences and evaluations in understanding frontline work. Narrative methods allowed for the study of normative reasoning employed by officials when making decisions (e.g. Maynard-Moody & Musheno, 2000; 2003). Moreover, the inductive approach involving storytelling and participant observation also allowed for the study of social dynamics, which have largely remained out of sight in the street-level bureaucracy literature. This study has shown that inductive methods are also needed to study the improvisational and unplanned acts of frontline officials. The use of such methods has brought insight in frontline behavior from the perspective of frontline officials themselves, allowing to go beyond top-down notions as the application of knowledge and rules, and discretion.

Secondly, this dissertation has offered a method to studying the effects of stereotypes in context. Sociological studies on frontline stereotyping have typically used interpretive methods, such as storytelling, in-depth interviewing and participant observations to study frontline officials' categorizations and experiences (e.g. Dubois, 2013, 2014; Maynard-Moody & Musheno, 2003; 2012). Public administration scholars, in their turn, have started studying frontline discrimination using experimental research designs (e.g. Andersen & Guul, 2016). The first branch of literature shows how frontline officials' decisions are intertwined with stereotypical reasoning, but do not allow for studying the effects of stereotypes on street-level judgment. The experimental designs, to the contrary, do allow for the study of such effects, but these do generally not offer insight in the context of stereotyping. This dissertation has used a policy-capturing design that allows for the study of the influence of stereotypes on frontline interpretation, while

also taking into consideration the context of making frontline evaluations. Respondents were asked to evaluate a fairly large number of hypothetical scenarios within an interview setting, which resembles real-life settings where officials do not make evaluations in a vacuum. The findings of the policy-capturing study, therefore, likely resemble real-life settings where officials' decisions are also made in comparison to other cases, and where officials typically try to 'block' their prejudices and to 'postpone' their first impressions, because they have learned to do so.

The combined use of the vignettes and interview data, furthermore, allowed me to gain insight in both officials' stereotypical reasoning as also the more implicit effects of status characteristics on officials' evaluations. Whereas the interviews primarily gave insight in the forms of stereotyping that are considered more acceptable, and the reasoning behind this, the policy-capturing study allowed for the study of forms of stereotyping that are more implicit and considered less acceptable. The interview that was part of the policy-capturing study showed how some officials indeed have lower expectations of lower educated entrepreneurs, but that these were typically associated with a more lenient treatment. In the interviews, officials never explicitly talked about evaluating lower status entrepreneurs stricter than higher status entrepreneurs. However, the policy-capturing study did in fact indicate that officials' evaluations of signals of lower educated entrepreneurs could be stricter than their evaluations of similar signals of higher educated entrepreneurs. Interviewing, thus, seems an appropriate method to study officials' stereotypical reasoning and their legitimizations, whereas a vignette approach is apt to examine the implicit 'force' of status characteristics on frontline evaluations.

6.6. Limitations of the study

This study also has some limitations which need to be discussed. A first limitation is the dissertation's focus on a particular type of street-level bureaucrat in only one context. Whereas the focus on frontline tax officials has been motivated by theoretical considerations, it also hampers generalizability to other contexts. It is highly likely that the findings are generalizable to frontline contexts where officials have considerable discretion and have been encouraged to rely on their own professional judgment in making decisions, which are information- and knowledge-intensive, and are surrounded by considerable rules and regulations. Whether the same findings can also be expected in a bureaucratic context which is less information driven, involves less expert knowledge, and where officials have less discretion, remains an empirical question. Furthermore, whereas this study has shown the importance of officials' reliance on colleagues in making decisions, the practice of inspection and decision making by tax officials is still highly individualized. Different findings may be expected in contexts where team work and peer review is facilitated or required by the organization. In such context, the implicit social dynamics in groups may play an even more important role.

Secondly, whereas the policy-capturing study tested the sociological double standards mechanism, the social character of this form of stereotyping is primarily in the focus on typical status characteristics, and not empirically assessed as such. The study looked specifically at citizen-clients' social class and level of education, which are characteristics that are typically associated with more generic cultural beliefs about people's overall competence. Lower educated people are viewed as generally less competent than higher educated people (e.g. Spence, 1973). Moreover, lower social classes are also perceived as less competent and in need of help (e.g. Harrits & Møller, 2014; Dubois, 2010). This study assessed whether these stereotypes also have their force in the official-entrepreneur encounter, by examining whether similar signals from these status groups are evaluated differently. How officials' stereotypical judgments are affected by

stereotypes existing in society at large is thus not empirically assessed, but assumed. The main contribution this particular study has made is proposing and testing the ‘stereotypes as frame mechanism’, which has hitherto scarcely been studied within the public administration literature on stereotyping.

Thirdly, a methodological limitation is chapter two’s use of data which were independently gathered by two researchers. The data of each case were collected with different purposes, and not with the aim to study social context. Although differences between organizational contexts, and even between countries may have affected the specific findings, the fact that we found that the social context has an impact in all constellations strengthens our more general claim that the social context of frontline decision making is an important topic of study.

Lastly, the null findings of the policy-capturing study could be the result of some methodological limitations. Whereas the qualitative study (chapter 4) indicated that officials may use differential standards, most of the hypothesized interactions (chapter 5) were non-significant. It could be that the qualitative interviews were not feasible to examine double standards, because this method does not allow for the control of the research context which is necessary to ensure the comparability of situations. In that regard, the method of interviewing is suitable for studying stereotypical beliefs, but less for studying indirect mechanisms. Another relevant factor could be the framing of the scenarios. Many respondents argued that some cues are of greater importance in more ‘negative’ situations where faults have been determined and a fine must be decided upon. Whereas some scholars would argue social desirability could have played a role, the question is whether officials also try to be non-biased in their daily judgments. The interview which was part of the policy-capturing study has shown that some respondents argued they learned to block their prejudices and postpone their first impressions in their actual work. Also, since this study did find a significant moderation effect that turned out more negative for lower status group than the higher status group, it is likely that the use of vignettes combining different cues partly mitigated the problem of strategic answering and allowed for the study of implicit stereotyping.

6.7. Recommendations for future research

Based on the findings and limitations of this study, I will also propose some avenues for future research. Firstly, some methodological recommendations for the study of frontline work and judgment can be made. This dissertation has proposed a bottom-up perspective to the study of frontline work, allowing for the study of officials’ own perspectives and experiences. To this end, this study used inductive research methods, such as storytelling and participant observations, which are suitable to study officials’ meaning making. This is not to say that more deductive research designs are not valuable. To the contrary, deductive qualitative and quantitative research designs could also be used within bottom-up approaches, and are valuable in constructing and testing theories, provided that these designs enable the study of context. This context could either be the explicit meaning making of officials, as their assessments and behavior driven by implicit meanings. In line with this study, future studies could combine designs looking at officials’ meaning making with studies looking at the implicit drivers of frontline behavior. Such studies would give insight in what officials find meaningful in their daily work, but also allow for studying the influence of ‘implicit factors’, such as stereotypes or routines, which could go beyond officials’ own meaningful understanding.

Secondly, this study has shown the added value of using different methods to study different aspects of stereotyping. Policy-capturing or vignette studies allow for the study of implicit ste-

reotyping in context, whereas qualitative methods allow for the study of officials' stereotyped reasoning and expectations. Future studies on frontline stereotyping could continue this effort by combining different methods, such as policy-capturing and traditional experiments with treatment and control groups, to study similar research questions. In doing this, the specific contributions of each method to the study of stereotyping could be assessed and compared.

The findings of this dissertation are furthermore of relevance for the literature on bureaucracy and bureaucratic uncertainty. This study has found that social dynamics, stereotypes and double standards could affect frontline decision making, leading to inconsistency and arbitrariness. This study has pointed to different 'levels' and 'kinds' of inconsistency. With regard to the levels of inconsistency, the study has shown how similar situations are evaluated differently for different status groups by the same official (chapter 5), but also to differences in decision making between officials and even organizational units (chapter 3). With regards to the kinds of inconsistency, differences between cases can be made based on the perceived relationship, but also based on citizen-clients' attributes, such as their social class or level of education. Some of the differences being made are seen as legitimate by officials, others are not, and some occur implicitly, and others more deliberately. It is not up to scientists to decide what kinds of inconsistency are acceptable or not, but to empirically assess how officials make differences between cases, when this is perceived by frontline officials and their managers as justified and when not, and how these practices relate to the bureaucratic notions of uncertainty and uncertainty reduction.

A recommendation for future research, then, is to more analytically distinguish between various levels of inconsistency, such as the individual level (within one person), the meso-level (between colleagues in the same organization) and the organizational level (between organizations). Moreover, future research could study the bases on which officials deliberately but also implicitly make differences between otherwise similar cases, and how this relates to the organizational policy's emphasis on either consistency or responsiveness. It could, for example, be studied whether officials who are encouraged to be responsive are more inclined to make differences based on characteristics of the interaction or social stereotypes than officials encouraged to be consistent.

Based on the findings of this study some recommendations for future research can also be offered to street-level bureaucracy scholars. The storytelling study focused on uncertainty experiences has shown that officials look for proof to substantiate their account (chapter 2). Also, the qualitative study focused on officials' evaluations of citizen-clients' trustworthiness has shown that signals do not have a conclusive meaning on their own, but that it is the whole picture that has already been formed that determines how signals are interpreted (chapter 4). Chapter four and five, subsequently studied and suggested how citizen-clients' status characteristics could serve as moderating contexts, influencing the interpretation of other signals. These findings point to the relevance of studying how officials construct images of citizen-clients, and the interpretation work that is part and parcel of collecting signals and proof. Future research, then, not only should focus on studying the kinds of signals officials gather, such as the status of someone's bookkeeping records, but also examining how the image that has up until then been formed affects the interpretation of such a signal. This study has focused on the effects of citizen-clients' status characteristics on that interpretation. Future studies could look at how other aspects of this image, such as demeanor, affect the interpretation of signals.

Another recommendation for street-level bureaucracy scholars is to study frontline officials' improvisation and the role of emotional labor in frontline behavior (e.g. Guy et al., 2008). This study has shown that on-the-spot judgments and the management of emotions are part and parcel of frontline work. This requires a bottom-up approach that goes beyond the notions of

discretion and application of rules and knowledge, to study officials' own accounts of their day-to-day work and judgments (e.g. Maynard-Moody & Musheno, 2003, 2012; Gofen, 2013). Such an approach furthermore enables the study of how frontline decision making is affected by social dynamics.

The street-level bureaucracy literature would furthermore benefit from comparative approaches to the study of frontline work (Hupe et al., 2015). Similar studies could be conducted in different bureaucratic contexts in order to develop a theory on for instance the relationship between stereotyping and the leeway that is officially granted to officials. It is assumable that, for example, a reliance on particularistic signals and double standards is more prevalent in bureaucratic contexts where officials have considerable room for interpretation and which emphasize responsiveness and profession, than in more rule-bound contexts emphasizing consistency. The same can be expected with regards to the experience of interpretation and action uncertainties, which may be more prevalent in the former than in the latter context. Such comparative studies would give a deeper insight in the bureaucratic conditions under which officials experience certain kinds of uncertainty and make particularistic judgments.

A last recommendation for future research on street-level bureaucracy is to study the other side of the bureaucratic interaction, that is, the citizen-clients. This study has shown how frontline officials sometimes negotiate with citizen-clients, and can be more responsive towards particular cases. Also, officials sometimes fear they may lose control over situations when they are in interaction with citizen-clients. This makes the study of the perspectives and strategies of citizen-clients when interacting with frontline officials a relevant topic of study. An interesting avenue for research may be to study whether and how citizen-clients' strategies affect decision-making outcomes, and the role of their background in the extent to which they have an influence.

Some avenues for future research can also be offered to the public management literature. Prior research has shown that rules could decrease the risk of inconsistency in decision making, but that it could also hamper officials to be responsive (Evans, 2016; Loyens & Maesschalk, 2010). This study has shown the importance of considering the social context of frontline decision making. Officials sometimes rely on their coworkers and managers to come to a judgment that is less subjective and is backed-up by others. Future research could more systematically study the ways in which uncertainties are dealt with socially, and look at the effects of such deliberation practices on the consistency of decisions across the organization. Furthermore, future studies could study how different organizational forms and cultures relate to officials' willingness to talk openly and reflect on cases and the decisions being made.

Moreover, future research could also particularly look at the role of the manager in dealing with both consistency and flexibility. Managers are traditionally seen as the agents of hierarchical control who are committed to implementing organizational policy and aimed at reducing frontline uncertainty (e.g. Evans, 2011). However, public organizations employ professionals or emphasize professional frontline judgment precisely because of a felt need for more flexibility (Evans, 2011; Groeneveld, 2016). Research has shown that public managers can have different roles and not only feel loyal to organizational goals, but also to professional frontline workers and citizen-clients (Evans, 2011; Noordegraaf & De Wit, 2012). Future research could extend this line of research, by studying public managers' roles in deliberation practices among frontline officials and in the creation of shared, informal norms and interpretations of cases (e.g. Groeneveld, 2016; Piore, 2011).

Recommendations for future research can also be offered to scholars interested in studying stereotyping by public officials. Both the interview study and policy-capturing study on front-

line stereotyping primarily focused on what and how characteristics of citizen-clients affect frontline decision making, and not on characteristics of frontline officials themselves. Future research could examine the effect of officials' background characteristics, such as tenure, level of education and ethnicity, on their use of stereotypes or double standards. Such a study would also offer valuable insights to the representative bureaucracy literature. Research on active representation studies whether and how officials from disadvantaged groups use their position actively to promote the interests of the group in society they came from (e.g. Meier & Bohte, 2001). Whereas this study suggests a variety of subjective judgment practices, future studies could examine whether and how officials' background characteristics align with their use of social typologies. Such a study would give insight in whether and how bureaucratic uncertainty reduction takes shape by power distributions existing in society at large.

Bibliography

- Aguinis, H., & Bradley, K.J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351-371.
- Aiman-Smith, L., Scullen, S.E., & Barr, S.H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, 5(4), 388-414.
- Alexander, J.C. (2005). Why cultural sociology is not 'idealist'? A reply to McLennan. *Theory, Culture & Society*, 22(6), 19-29.
- Andersen, S.C., & Guul, T.S. (2016, January). *Minority discrimination at the front line. Combined survey and field experimental evidence*. Paper prepared for presentation at the Annual Meeting of the Southern Political Science Association, San Juan, Puerto Rico.
- Barber, B. (1983). *The logic and limits of trust*. New Brunswick, NJ: Rutgers University Press.
- Bartels, K.P.R. (2013). Public encounters: The history and future of face-to-face contact between public professionals and citizens. *Public Administration*, 91(2), 469-483.
- Bartels, K.P.R. (2014). Communicative capacity: The added value of public encounters for participatory democracy. *The American Review of Public Administration*, 44(6), 656-674.
- Behn, R.D. (2001). *Rethinking democratic accountability*. Washington, DC: Brookings Institution Press.
- Belastingdienst (2016). *Controleaanpak Belastingdienst (CAB): De CAB en zijn modellen toegepast in toezicht*. Retrieved from: https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/themaoverstijgend/brochures_en_publicaties/controleaanpak_belastingdienst
- Berger, J., Cohen, B.P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 37(3), 241-255.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66(1), 5-20.
- Biernat, M., Manis, M., & Nelson, T.E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60(4), 485-499.
- Bovens, M. (1998). *The quest for responsibility: Accountability and citizenship in complex organizations*. Cambridge, UK: Cambridge University Press.
- Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174-184.
- Brehm, J., & Gates, S. (1997). *Working, shirking, and sabotage: Bureaucratic response to a democratic public*. Ann Arbor, Michigan: University of Michigan Press.

Bibliography

- Brodkin, E.Z. (2011). Policy work: Street-level organizations under new managerialism. *Journal of Public Administration Research and Theory*, 21(2), i253-1277.
- Brodkin, E.Z. (2015). The inside story: Street-level research in the US and beyond. In P. Hupe, M. Hill, & A. Buffat (Eds.), *Understanding street-level bureaucracy* (pp. 25-42). Bristol: Policy Press.
- Bruhn, A., & Ekström, M. (2017). Towards a multi-level approach on frontline interactions in the public sector: Institutional transformations and the dynamics of real-time interactions. *Social Policy & Administration*, 51(1), 195-215.
- Buffat, A. (2013). Street-level bureaucracy and e-government. *Public Management Review*, 17(1), 149-161.
- Clarke, J., & Newman, J. (1997). *The managerial state: Power, politics and ideology in the remaking of social welfare*. London: Sage.
- Clayton, R., & Pontusson, J. (1998). Welfare-state retrenchment revisited: Entitlement cuts, public sector restructuring, and inegalitarian trends in advanced capitalist societies. *World Politics*, 51(1), 67-98.
- Cohen, N., & Gershgoren, S. (2016). The incentives of street-level bureaucrats and inequality in tax assessments. *Administration & Society*, 48(3), 267-289.
- Cooksey, R.W. (1996). *Judgment analysis: Theory, methods and applications*. San Diego, CA: Academic Press.
- Correll, S.J., & Benard, S. (2006). Biased estimators? Comparing status and statistical theories of gender discrimination. In S.R. Thye, & E.J. Lawler (Eds.), *Advances in Group Processes* (pp. 89-116). Bingley, UK: Emerald Group Publishing Limited.
- Correll, S.J., & Ridgeway, C.L. (2003). Expectation states theory. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 29-51). New York: Kluwer Academic/Plenum Publishers.
- Das, T.K., & Teng, B. (2004). The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology*, 19(1), 85-116.
- Davis, K.C. (1969). *Discretionary justice: A preliminary inquiry*. Baton Rouge, LA: Louisiana State University Press.
- Deusch, F.M., Zalski, C.M., & Clark, M.E. (1986). Is there a double standard of aging? *Journal of Applied Social Psychology*, 16(9), 771-785.
- Downs, A. (1966). *Bureaucratic structure and decisionmaking*. Santa Monica, California: The Rand Corporation.
- Dubois, V. (2010). *The bureaucrat and the poor: Encounters in French welfare offices*. Burlington, VT: Ashgate.
- Dubois, V. (2013). The functions of bureaucratic routines in a changing welfare state: On interactions with recipients in French welfare offices. In Ph. Sander mann (Ed.), *The end of welfare as we know it? Continuity and change in western welfare practices* (pp. 127-136). Opladen, Berlin & Toronto: Barbara Budrich Publishers.
- Dubois, V. (2014). The state, legal rigor, and the poor: The daily practice of welfare control. *Social Analysis*, 58(3), 38-55.
- Epp, C.R., Maynard-Moody, S., & Haider-Markel, D.P. (2014). *How police stops define race and citizenship*. Chicago & London: The University of Chicago Press.
- Erickson, F. (1975). Gatekeeping and the melting pot: Interaction in counseling encounters. *Harvard Educational Review*, 45(1), 44-70.
- Evans, T. (2011). Professionals, managers and discretion: Critiquing street-level bureaucracy. *The British Journal of Social Work*, 41(2): 368-386.
- Evans, T. (2015). Professionals and discretion in street-level bureaucracy. In P. Hupe, M. Hill,

- & A. Buffat (Eds.), *Understanding Street-Level Bureaucracy* (pp. 279-293). Bristol & Chicago, IL: Policy Press.
- Evans, T. (2016). Professional discretion in welfare services: Beyond street-level bureaucracy. London & New York: Routledge.
- Fassin, D. (2013). The precarious truth of asylum. *Public Culture*, 25(1), 39-63.
- Fineman, S. (1998). Street-level bureaucrats and the social construction of environmental control. *Organization Studies*, 19(6), 953-974.
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 26(1), 21-42.
- Gajduschek, G. (2003). Bureaucracy: Is it efficient? Is it not? Is that the question? Uncertainty reduction: An ignored element of bureaucratic rationality. *Administration & Society*, 34(6), 700-723.
- Gambetta, D., & Hamill, H. (2005). *Streetwise: How taxi drivers establish their customers' trustworthiness*. New York: Russell Sage Foundation.
- Gofen, A. (2013). Mind the gap: Dimensions and influence of street-level divergence. *Journal of Public Administration Research and Theory*, 24(2), 473-493.
- Goodsell, C.T. (1981). *The public encounter: Where state and citizen meet*. Bloomington, US: Indiana University Press.
- Gribnau, H. (2007). Soft law and taxation: The case of the Netherlands. *Legisprudence*, 1(3), 291-326.
- Grimmelikhuijsen, S., Jilke, S., Olsen, A.L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, 77(1), 45-56.
- Groeneveld, S. (2016). *Het belang van bureaucratie. Omgaan met ambivalentie in publiek management*. Inaugural lecture (2016, May 27). Leiden: Leiden University.
- Groeneveld, S., & Van de Walle, S. (2011). Introduction. In S. Groeneveld, & S. Van de Walle (Eds.), *New steering concepts in public management* (pp. 1-8). UK, Bingley: Emerald Group Publishing Limited.
- Guy, M.E., Newman, M.A., & Mastracci, S.H. (2008). *Emotional labor: Putting the service in public service*. Armonk & London: M.E. Sharpe.
- Handelman, D. (1978). Bureaucratic interpretation: The perception of child abuse in urban Newfoundland. In D. Handelman, & E. Leyton (Eds.), *Bureaucracy and world view: Studies in the logic of official interpretation* (pp. 15-69). Canada: University of Toronto Press.
- Harris, D.A. (1999). The stories, the statistics and the law: Why 'driving while black' matters. *University of Minnesota Law Review*, 84(2), 265-326.
- Harrits, G.S. (2011). More than method? A discussion of paradigm differences within mixed methods research. *Journal of mixed methods research*, 5(2), 150-166.
- Harrits, G.S., & Møller, M.Ø. (2014). Prevention at the front line: How home nurses, pedagogues, and teachers transform public worry into decisions on special efforts. *Public Management Review*, 16(4), 447-480.
- Hatmaker, D.M., Park, H.H., & Rethemeyer, K. (2011). Learning the ropes: Communities of practice and social networks in the public sector. *International Public Management Journal*, 14(4), 395-419.
- Hoag, C. (2011). Assembling partial perspectives: Thoughts on the anthropology of bureaucracy. *Political and Legal Anthropology Review*, 34(1), 81-94.
- Hupe, P., & Hill, M. (2007). Street-level bureaucracy and public accountability. *Public Administration*, 85(2), 279-299.

Bibliography

- Hupe, P., Hill, M., & Buffat, A. (Eds.) (2015). *Understanding street-level bureaucracy*. Bristol & Chicago, IL: Policy Press.
- Jasso, G., & Webster, M. Jr. (1997). Double standards in just earnings for male and female workers. *Social Psychology Quarterly*, 60(1), 66-78.
- Johnson, R.R., & Morgan, M.A. (2013). Suspicion formation among police officers: An international literature review. *Criminal Justice Studies*, 26(1), 99-114
- Jones, B.D. (2001). *Politics and the architecture of choice: Bounded rationality and governance*. Chicago: The University of Chicago Press.
- Jorna, F., & Wagenaar, P. (2007). The iron cage strengthened? Discretion and digital discipline. *Public Administration*, 85(1), 189-214.
- Karren, R.J., & Barringer, M.W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, 5(4), 337-361.
- Katz, D., Gutek, B.A., Kahn, R.L., & Barton, E. (1975). *Bureaucratic encounters: A pilot study in the evaluation of government services*. Ann Arbor, Michigan: The University of Michigan.
- Keiser, L. R. (2010). Understanding street-level bureaucrats' decision making: Determining eligibility in the Social Security Disability Program. *Public Administration Review*, 70(2), 247-257.
- Kirchler, E. (2007). *The economic psychology of tax behaviour*. Cambridge: Cambridge University Press.
- Kravel-Tovi, M. (2012). Rite of passing: Bureaucratic encounters, dramaturgy and Jewish conversion in Israel. *American Ethnologist*, 39(2), 371-388.
- Leviner, S. (2009). A new era of tax enforcement: From 'big stick' to responsive regulation. *University of Michigan Journal of Law Reform*, 42(2), 381-429.
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Loyens, K. (2012). *Integrity secured. Understanding ethical decision making among street-level bureaucrats in the Belgian Labor Inspection and Federal Police*. (Doctoral dissertation, KU Leuven, Leuven, Belgium). Retrieved from <https://lirias.kuleuven.be/handle/123456789/347145>
- Loyens, K., & Maesschalck, J. (2010). Toward a theoretical framework for ethical decision making of street-level bureaucracy: Existing models reconsidered. *Administration & Society*, 42(1), 66-100.
- Maroulis, S. (2015). The role of social network structure in street-level innovation. *The American Review of Public Administration*, DOI: 10.1177/0275074015611745.
- Mascini, P., & Van Wijk, E. (2009). Responsive regulation at the Dutch food and consumer product safety authority: An empirical assessment of assumptions underlying the theory. *Regulation & Governance*, 3(1), 27-47.
- May, P.J., & Winter, S.C. (2009). Politicians, managers, and street-level bureaucrats: Influences on policy implementation. *Journal of Public Administration Research and Theory*, 19(3), 453-476.
- May, P.J., & Winter, S.C. (2011). Regulatory enforcement styles and compliance. In C. Parker, & V.L. Nielsen (Eds.), *Explaining regulatory compliance: Business responses to regulation* (pp. 222-244). Cheltenham, UK & Northampton, MA, USA: Edward Elgar Publishing.
- Maynard-Moody, S., & Musheno, M. (2000). State agent or citizen agent: Two narratives of discretion. *Journal of Public Administration Research and Theory*, 10(2), 329-358.
- Maynard-Moody, S., & Musheno, M. (2003). *Cops, teachers, counselors: Stories from the front lines of public service*. Ann Arbor: The University of Michigan Press.

- Maynard-Moody, S., & Musheno, M. (2012). Social equities and inequities in practice: Street-level workers as agents and pragmatists. *Public Administration Review*, 72(s1) S16-S23.
- McCallin, A. (2001). Inter-disciplinary practice - A matter of teamwork: An integrated literature review. *Journal of Clinical Nursing*, 10(4), 419-428.
- Meier, K.J., & Bohte, J. (2001). Structure and discretion: Missing links in representative bureaucracy. *Journal of Public Administration Research and Theory*, 11(4), 455-470.
- Mennerick, L.A. (1974). Client typologies: A method of coping with conflict in the service worker-client relationship. *Work and Occupations*, 1(4), 396-418.
- Merton, R.K. (1940). Bureaucratic structure and personality. *Social Forces*, 18(4), 560-568.
- Møller, M.Ø., & Harrits, G.S. (2013). Constructing at-risk target groups. *Critical Policy Studies*, 7(2), 155-176
- Moyson, S., Raaphorst, N., Groeneveld, S., & Van de Walle, S. (2017). Organizational socialization in public administration research: A systematic review and directions for future research. *The American Review of Public Administration*. DOI: 10.1177/0275074017696160
- Newman, J. (1998). Managerialism and social welfare. In G. Hughes & G. Lewis (Eds.), *Unsettling welfare: The reconstruction of social policy* (pp. 333-374). London: Routledge.
- Nielsen, V. L. (2007). Differential treatment and communicative interactions: Why the character of social interaction is important. *Law and Policy*, 29(2), 257-283.
- Nielsen, V.L. (2015). Law enforcement behaviour of regulatory inspectors. In P. Hupe, M. Hill, & A. Buffat (Eds.), *Understanding street-level bureaucracy* (pp. 115-131). Bristol & Chicago, IL: Policy Press.
- Noordegraaf, M. (2011). Risky business: How professionals and professional fields (must) deal with organizational issues. *Organizational Studies*, 32(10), 1349-1371.
- Noordegraaf, M. (2016). Reconfiguring professional work: Changing forms of professionalism in public services. *Administration & Society*, 48(7), 783-810.
- Noordegraaf, M., & De Wit, B. (2012). How management pressures affect managerial relations and loyalties in education. *Public Administration*, 90(4), 957-973.
- Nooteboom, B. (1996). Trust, opportunism and governance: A process and control model. *Organization Studies*, 17(6), 985-1010.
- Oberfield, Z.W. (2014). *Becoming bureaucrats: Socialization at the front lines of government service*. Pennsylvania: University of Pennsylvania Press.
- Peters, B.G. (2004). Governance and public bureaucracy: New forms of democracy or new forms of control? *Asia Pacific Journal of Public Administration*, 26(1), 3-15.
- Piore, M.J. (2011). Beyond markets: Sociology, street-level bureaucracy, and the management of the public sector. *Regulation & Governance*, 5(1), 145-164.
- Pollitt, C. (2013). *Context in public policy and management: The missing link?* Cheltenham: Edward Elgar Publishing.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. Oxford: Oxford University Press.
- Prottas, J.M. (1979). *People-processing: The street-level bureaucrat in public service bureaucracies*. Lexington: Lexington Books.
- Rice, D. (2013). Street-level bureaucracy and the welfare state. *Administration & Society*, 45(9), 1038-1062.
- Ridgeway, C. (1991). The social construction of status value: Gender and other nominal characteristics. *Social Forces*, 70(2), 367-386.
- Ridgeway, C.L. (2014). Why status matters for inequality. *American Sociological Review*, 79(1), 1-16.

Bibliography

- Rutz, S., Mathew, D., Robben, P., & De Bont, A. (2015). Enhancing responsiveness and consistency: Comparing the collective use of discretion and discretionary room at inspectorates in England and the Netherlands. *Regulation & Governance*, DOI: 10.1111/rego.12101
- Sandfort, J. (1999). The structural impediments to human service collaboration: Examining welfare reform at the front lines. *Social Service Review*, 73(3), 314-339.
- Sandfort, J.R. (2000). Moving beyond discretion and outcomes: Examining public management from the front lines of the welfare system. *Journal of Public Administration Research and Theory*, 10(4), 729-756.
- Savi, R. (2014). Public policy-making in a time of cutbacks: The role of the street-level bureaucrats in Estonia. *Administrative Culture*, 15(1), 100-119.
- Schott, C., Van Kleef, D., & Noordegraaf, M. (2016). Confused professionals? Capacities to cope with pressures on professional work. *Public Management Review*, 18(4), 583-610.
- Schram, S.F., Soss, J., Fording, R.C., & Houser, L. (2009). Deciding to discipline: Race, choice, and punishment at the frontlines of welfare reform. *American Sociological Review*, 74(3), 398-422.
- Schuppan, T. (2015). Service workers on the electronic leash? Street-level bureaucrats in emerging information and communication technology work contexts. In P. Hupe, M. Hill, & A. Buffat (Eds.), *Understanding street-level bureaucracy* (pp. 243-260). Bristol & Chicago, IL: Policy Press.
- Siciliano, M.D. (2015). Professional networks and street-level performance: How public school teachers' advice networks influence student performance. *The American Review of Public Administration*, 47(1), 79-101.
- Simmel, G. (1908). *Soziologie, Untersuchungen über die Formen der Vergesellschaftung*. Leipzig, Germany: Verlag von Duncker und Humblot.
- Simon, H.A. (1976). *Administrative behavior: A study of decision-making processes in administrative organization* (3th ed.). New York: Macmillan.
- Snellen, I. (1998). Street level bureaucracy in an information age. In E. Snellen, & W. van de Donk (Eds.), *Public administration in an information age: A handbook* (pp. 497-505). Amsterdam: IOS Press.
- Snijders, T.A.B., & Berkhof, J. (2007). Diagnostic checks for multilevel models. In J. de Leeuw, & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 139-173). New York: Springer.
- Soss, J., Fording, R.C., & Schram, S.F. (2011a). The organization of discipline: From performance management to perversity and punishment. *Journal of Public Administration Research and Theory*, 21(suppl. 2), i203-i232.
- Soss, J., Fording, R.C., & Schram, S.F. (2011b). *Disciplining the poor: Neoliberal paternalism and the persistent power of race*. Chicago, IL: University of Chicago Press.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.
- Stewart, T.R. (1988). Judgment analysis: Procedures. In B. Brehmer, & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 41-74). Amsterdam: North-Holland Elsevier.
- Stiglitz, J.E. (1975). The theory of "screening", education and the distribution of income. *The American Economic Review*, 65(3), 283-300.
- Stivers, C. (1994). The listening bureaucrat: Responsiveness in public administration. *Public Administration Review*, 54(4), 364-369.
- Stone, D. (1984). *The disabled state*. Philadelphia, PA: Temple University Press.
- Stroshine, M., Alpert, G., & Dunham, R. (2008). The influence of "working rules" on police suspicion and discretionary decision making. *Police Quarterly*, 11(3), 315-337.
- Thacher, D., & Rein, M. (2004). Managing value conflict in public policy. *Governance*, 17(4),

- Thompson, J.D. (2003). *Organizations in action: Social science bases of administrative theory*. New Brunswick and London: Transaction Publishers. (Original work published in 1967).
- Tummers, L. (2016). The relationship between coping and job performance. *Journal of Public Administration Research and Theory*, 27(1), 150-162.
- Tummers, L.L.G., Bekkers, V., Vink, E., & Musheno, M. (2015). Coping during public service delivery: A conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory*, 25(4), 1099-1126.
- Van der Aa, P., & Van Berkel, R. (2015). Fulfilling the promise of professionalism in street-level practice. In P. Hupe, M. Hill, & A. Buffat (Eds.), *Understanding street-level bureaucracy* (pp. 263-278). Bristol & Chicago, IL: Policy Press.
- Van Kleef, D. (2016). *Changing the nature of the beast: How organizational socialization contributes to the development of the organizational role identity of Dutch veterinary inspectors*. (Doctoral dissertation, Leiden University, Leiden, The Netherlands). Retrieved from <https://openaccess.leidenuniv.nl/handle/1887/39944>
- Vigoda, E. (2002). From responsiveness to collaboration: Governance, citizens, and the next generation of public administration. *Public Administration Review*, 62(5), 527-540.
- Vinzant, J.C., & Crothers, L. (1998). *Street-level leadership: Discretion and legitimacy in front-line public service*. Washington D.C.: Georgetown University Press.
- Vis, B., Van Kersbergen, K., & Hylands, T. (2011). To what extent did the financial crisis intensify the pressure to reform the welfare state? *Social Policy & Administration*, 45(4), 338-353.
- Wagenaar, H. (2004). "Knowing" the rules: Administrative work as practice. *Public Administration Review*, 64(6), 643-656.
- Wagner, D.G., & Berger, J. (1993). Status characteristics theory: The growth of a program. In J. Berger, & M. Jr. Zelditch (Eds.), *Theoretical research programs: Studies in the growth of theory* (pp. 23-63). Stanford, California: Stanford University Press.
- Weber, M. (1947). *The theory of social and economic organization*. (A.M. Henderson & T. Parsons, Trans.). New York: Oxford University Press. (Original work published in 1922)
- Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. Berkeley: University of California Press.
- Yang, K. (2005). Public administrators' trust in citizens: A missing link in citizen involvement efforts. *Public Administration Review*, 65(3), 273-285.

Appendices

Appendix 1 – Coding table

Kinds of uncertainty in context	When do they occur?
Information uncertainty	
Finding proof	Knowing or feeling something is the case but not having (enough) evidence (4)
Interpretation uncertainty	
Experiencing dilemmas	Experiencing tension between task as tax official, feelings of empathy and ideas of justice (8)
Determining the right decision	Working with vague or too generic rules and regulation (4)
Action uncertainty	
Negotiating with clients	Not knowing what other party will bring to table or accept and whether official gives away too much (3)
Responding to client emotions and private situations	Encountering... ...influence of clients' private situation on encounter (4) ...influence of clients' emotions on encounter (5) ...a client infringing on official's private life (1)
Encountering deviations from normalcy	Encountering... ...foot-dragging by clients (3) ...complex, messy or absent bookkeeping records (3) ...unexpected reaction of client (2)

Appendix 2 – Interview grid

	Bad intentions	Good intentions
Incompetent	- Story 1 -	- Story 2 -
Competent	- Story 3 -	- Story 4 -

Appendix 3 – Scenario example



The undermentioned inspection is part of a random sample. The entrepreneur does not have an advisor. Read the description and answer the statements for the described inspection.

You get the task to conduct an inspection at an entrepreneur with a clothing store. It's an one-man business, situated in the street you see in the picture. Your preparation did not yield any particularities. The respective entrepreneur avoids contact with you during the introductory meeting. During the inspection you notice that some invoices are missing from the records. You have the impression that the entrepreneur is lower educated.

	1 Totally disagree	2 Disagree	3 Somewhat disagree	4 Neither disagree, nor agree	5 Somewhat agree	6 Agree	7 Totally agree
It seems fine here	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think the entrepreneur can be trusted	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would more critically look at this entrepreneur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 4 – Operationalization

Dependent variables – items (7-point Likert scale: totally disagree – totally agree)
Trust evaluation I think the entrepreneur can be trusted
Overall evaluation It seems fine here
Intended behavior I would more critically look at this entrepreneur

Cues – behavioral statements & pictures
Quality of bookkeeping
1. You notice that hardly any records are kept
2. You notice that some invoices are missing from the records
3. You notice that the invoices in the records are numbered consecutively and continuously
Quality of interaction
1. The entrepreneur avoids contact with you
2. The entrepreneur talks around your questions
3. The entrepreneur answers your questions to the point
Level of education
1. You've the impression that the entrepreneur is lower educated
2. You've the impression that the entrepreneur is higher educated
Social class *
1. Photo 1, 2, 3 & 4
2. Photo 5, 6, 7 & 8

* Photo 4 and 8 have been downloaded from the website Flickr and are royalty free. The other photos have been bought at a website that allows use for non-commercial purposes.



Appendices



Source photo 4: Flickr, made by FaceMePLS

Source photo 8: Flickr, made by Stipo Team for Urban Development

Appendix 5 – Correlation matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1: Appears okay	-								
V2: Trust	,812**	-							
V3: More critical scrutinization	,794**	,724**	-						
V4: Dummy missing invoices	,211**	,136**	,228**	-					
V5: Dummy invoices in order	,622**	,451**	,612**	,500**	-				
V6: Dummy dodge around question	,146**	,197**	,140**	,000	,000	-			
V7 Dummy to the point	,265**	,348**	,233**	,000	,000	,500**	-		
V8: Level of education	,019	-,034	-,007	,000	,000	,000	,000	-	
V9: Social class	-,015	-,006	,020	,000	,000	,000	,000	,000	-

** p < 0.01.

Appendix 6 – Interview procedure

Step 1 – introduction and background questions

Introduction

- a) Introducing myself and general topic of research
- b) Guaranteeing anonymity of data processing and confidentiality
- c) Explanation of procedure

Background questions

- a) When started as tax official? How?
- b) What kind of job before that?

Instructions given

- a) The scenarios describe audits. Although they resemble real audits, they are different because there is less information. We believe that inspectors are able to make assessments based on these scenarios. The scenarios look alike, but are different. Please read them carefully and look at the pictures.
- b) Because there is only concise information, we don't ask you to make a final judgment. It's rather a provisional assessment based on your first impression/feeling. We know there are other aspects you would commonly further investigate that could shed a whole different light on the case. We are not interested in that. Only take the mentioned information into consideration.
- c) We want to emphasize that we are really interested in your first impression, and not in what other people might expect, or in what you think you should do. We're looking for honest answers. We're not testing whether you do something good or wrong in this research.
- d) Please fill out the scenarios yourself. We can discuss possible questions or doubts afterwards. If you doubt about something, try to fill out the questions based on your own impression. Halfway, we'll stop for 5 minutes and I'll ask you some background questions.
- e) Try not to think too long before giving your answers; we're interested in your first impression.

Step 2 – first 20 scenarios

Researcher distanced herself, and made notes on:

- a) Atmosphere of interview (open/closed; signals of fatigue)
- b) Time respondents took to fill out first 20 scenarios
- c) Questions and remarks respondent had (were only answered and discussed during reflection)

Step 3 – background questions (around 5 minutes)

Background questions

- a) Function? Specialization?
- b) What kind of taxes?
- c) Projects?

Step 4 – last 20 scenarios

Researcher distanced herself, and made notes on:

- a) Atmosphere of interview (open/closed; signals of fatigue)
- b) Time respondents took to fill out last 20 scenarios
- c) Questions and remarks respondent had (were only answered and discussed during reflection)

Step 5 – reflection and disclosing more about study

Reflection

- a) How is evaluation task experienced? Difficult/easy?
 - i. Researcher made notes on how respondent interpreted certain cues/questions.

Disclosing of cues

- a) How to rank-order these cues so it reflects the importance these aspects play in getting a first impression in still uncertain situations (as described in scenarios)?
 - i. Researcher made notes when sensing possible desirable answering to scenarios.

Step 6 – small questionnaire and wrapping up

Small questionnaire

a) Propensity to trust; Level of education; Country of origin

Wrapping up

a) Thank you and small thank you gift

b) Informing about presentation of findings

Dutch summary

Onzekerheid in Bureaucratie

Probleemstelling van dit proefschrift

Binnen publieke organisaties wordt van oudsher belang gehecht aan de reductie van onzekerheid. Strikte regels, een duidelijke hiërarchie, en een onpersoonlijke houding van ambtenaren zijn de centrale kenmerken van het Weberiaanse ideaaltype van bureaucratie (Weber, 1978). Dit leidt tot de controleerbaarheid van processen en van de output van bureaucratisch handelen. Tegelijkertijd krijgen ambtenaren binnen nieuwere horizontale en interactieve vormen van bestuur juist méér ruimte om beslissingen te nemen. Hierbij wordt de ambtenaar niet slechts geacht om het voorgeschreven beleid uit te voeren, maar om besluiten te nemen op basis van professionele inschattingen. Discretionaire bevoegdheden worden binnen deze horizontale vormen van sturing niet alleen gezien als noodzakelijk, maar ook als waardevol om tot een weloverwogen besluit te komen.

Onzekerheidsreductie ligt ten grondslag aan verschillende theoretische benaderingen die worden bestudeerd binnen de bestuurskundige literatuur. Veelal gaat het om een *top-down* benadering waarbij onder andere is bestudeerd in hoeverre ambtenaren rationeel besluiten nemen, hoe beleid zoals geformuleerd van bovenaf wordt uitgevoerd, en of en hoe ambtenaren verantwoording afleggen. Hierbij wordt kortom vooral gekeken naar het gebrek aan *controle van* de menselijke factor als bron van onzekerheid in publieke organisaties. De literatuur over *street-level bureaucracy* (Lipsky, 1980) benadrukt de onzekerheid die inherent is aan de discretionaire bevoegdheden van *street-level bureaucrats*. Desondanks wordt de onzekerheid die onderdeel is van beslissingsbevoegdheid van deze contactambtenaren voornamelijk als gegeven beschouwd. Hoe contactambtenaren deze onzekerheden zelf ervaren en hoe zij onzekerheid proberen te reduceren is een onderbelicht thema binnen de *street-level bureaucracy* literatuur.

In een context waar contactambtenaren meer speelruimte hebben en worden geacht in te spelen op situationele omstandigheden, is het waarschijnlijk dat sociale interacties en particularistische aspecten een grotere rol gaan spelen in besluitvorming. Om inzicht te krijgen in onzekerheid en onzekerheidsreductie in een dergelijke bureaucratische context, is een *bottom-up* benadering vereist die oog heeft voor de ervaringen en interpretaties van ambtenaren zelf. In dit proefschrift heb ik me daarop gericht, geleid door de volgende centrale onderzoeksvraag:

Welke onzekerheden ervaren contactambtenaren – werkend binnen een bureaucratische context waar discretionaire bevoegdheden zijn uitgebreid – en hoe proberen ze deze onzekerheden te reduceren?

Deze centrale vraagstelling is opgedeeld in vier sub-vragen die met behulp van verschillende theoretische benaderingen en methoden zijn bestudeerd.

Setting van het onderzoek

Het onderzoek is grotendeels uitgevoerd binnen de Nederlandse Belastingdienst. De ambtenaren waar dit proefschrift voornamelijk naar heeft gekeken, hebben als taak de belastingaangiften en administraties van ondernemers uit het segment Midden- en Kleinbedrijf (MKB) te controleren, en stellen daartoe boekenonderzoeken in bij ondernemers. De Belastingdienst kan gekarakteriseerd worden als een typisch bureaucratische organisatie, waar ambtenaren werken in hiërarchische relaties en werkprocessen die in grote mate gestandaardiseerd zijn door middel van digitale systemen. Tegelijkertijd hebben de ambtenaren die werken voor het MKB-segment meer ruimte gekregen om beslissingen te nemen op basis van hun eigen professionele inschattingen.

In een van de deelonderzoeken is zowel data gebruikt die verzameld is binnen de Nederlandse Belastingdienst als data die gegenereerd is binnen de Belgische Arbeidsinspectie (dit laatste als onderdeel van het proefschrift van Loyens, 2012). Het combineren van deze twee casussen heeft ons in staat gesteld om te onderzoeken hoe verschillende sociale constellaties – interacties met burgers, maar ook verschillende teamconstructies – impact hebben op besluitvorming door ambtenaren.

Kernbevindingen

Het eerste empirische hoofdstuk van dit proefschrift (hoofdstuk 2) heeft onzekerheidservaringen van ambtenaren onderzocht door middel van de *storytelling* methode binnen semigestructureerde interviews. Mijn analyse van de verhalen van ambtenaren laat zien dat zij niet alleen informatie- en interpretatieonzekerheden ervaren, maar ook *handelingsonzekerheden*. Naast een tekort aan bewijs (informatieonzekerheid), of twijfel over besluitvormingsstandaarden (interpretatieonzekerheid), ervaren ambtenaren onzekerheid over hoe zij controle (terug) kunnen krijgen over situaties in interactie met ondernemers (actieonzekerheid).

In het geval van een tekort aan bewijs, tasten ambtenaren niet in het duister over wat er gaande is. Ambtenaren hebben integendeel juist een sterk idee over wat er aan de hand is, maar hebben niet genoeg bewijs om het standpunt te ondersteunen. In een dergelijk geval zetten ambtenaren meer controlemiddelen in *of* accepteren ze dat ze hun standpunt moeten aanpassen. Bij interpretatieonzekerheid ervaren ambtenaren twijfel over standaarden om situaties mee te beoordelen. Dit gaat over vage richtlijnen die ambtenaren weinig houvast bieden of over spanningen tussen regels en wetten enerzijds en gevoelens van empathie en rechtvaardigheid anderzijds. Onderhandelingen met ondernemers en onverwachte reacties van ondernemers maken ambtenaren alert, omdat zij *on-the-spot* moeten reageren. Bij deze handelingsonzekerheid hebben zij het gevoel dat hun *on-the-spot* acties zwaar wegen op het verloop en de uitkomst van een interactie. Deze bevinding benadrukt het belang van improvisatie voor het dagelijkse werk van ambtenaren.

Het tweede empirische hoofdstuk (hoofdstuk 3) heeft onderzocht hoe de sociale context van ambtenaren invloed heeft op hun besluitvorming. Er is in het bijzonder gekeken naar hoe sociale dynamieken binnen verschillende sociale constellaties – interacties tussen ambtenaren en ondernemers, en interacties tussen collega's of professionals in teams – een impact hebben op besluitvorming door ambtenaren. Om verschillende sociale constellaties te kunnen onder-

zoeken is zowel interviewdata van de Nederlandse Belastingdienst als ook interview- en observatiedata van de Belgische Arbeidsinspectie geanalyseerd. De studie wijst uit dat ambtenaren collega's betrekken in oordeelsvorming wanneer zij minder subjectieve besluiten willen nemen. Ambtenaren hebben hierbij het idee dat een collega hen zou kunnen wijzen op een eventuele tunnelvisie over de ondernemer. Tegelijkertijd laat de analyse ook zien dat *impliciete* sociale dynamieken in de vorm van ervaren *sociale druk* of een *sterke sociale band* een rol kunnen spelen in besluitvorming, en soms een grotere rol spelen dan inhoudelijke overwegingen.

Binnen het derde empirische deelonderzoek (hoofdstuk 4) is onderzocht naar welke signalen ambtenaren kijken om de betrouwbaarheid van ondernemers vast te stellen, en is geëxploreerd hoe sociale typologiën invloed kunnen hebben op de interpretatie van signalen. Er is hierbij gebruik gemaakt van semigestructureerde interviews met daarin een *storytelling* onderdeel, om inzicht te krijgen in hoe ambtenaren zich een beeld vormen van ondernemers. De studie heeft uitgewezen dat belastingambtenaren niet alleen kijken naar signalen gerelateerd aan de boekhouding van ondernemers, maar dat er ook wordt gekeken naar het gedrag van ondernemers, hun privésituaties en naar achtergrondkenmerken zoals sociale klasse, opleidingsniveau en etniciteit. De analyse heeft bovendien laten zien dat vergelijkbare signalen anders geïnterpreteerd kunnen worden voor ondernemers uit verschillende sociale groepen.

Vervolgens is in het laatste empirische hoofdstuk (hoofdstuk 5) door middel van de *policy-capturing* methode getest of ambtenaren vergelijkbare situaties anders beoordelen voor ondernemers met verschillende opleidingsniveaus en uit verschillende sociale klassen. In een kwalitatieve interviewsetting hebben respondenten tientallen vignetten geëvalueerd, in termen van een algemene inschatting van de casus, de betrouwbaarheid van de ondernemer en de neiging om de casus meer kritisch te onderzoeken. Deze vignetten bestonden uit hypothetische, maar realistische scenario's waarin boekenonderzoeken zijn nagebootst. De bevindingen wijzen uit dat het merendeel van de hypothesen niet bevestigd kan worden. De meeste signalen van kwaliteit van boekhouding en kwaliteit van interactie worden niet significant anders beoordeeld voor lagere statusgroepen dan voor hogere statusgroepen. Bevindingen suggereren echter ook dat, voor ambtenaren, statuskenmerken van ondernemers inderdaad als frames kunnen fungeren die hun interpretatie van andere signalen beïnvloeden. Wanneer een laag opgeleide ondernemer om een vraag heen draait wordt dit iets negatiever geëvalueerd dan wanneer een hoog opgeleide ondernemer dit doet. Het ontwijken van contact, daarentegen, wordt iets negatiever beoordeeld voor de hoog opgeleide ondernemer dan voor de laag opgeleide ondernemer.

Algemene conclusie

Er kan geconcludeerd worden dat ambtenaren verschillende soorten onzekerheden ervaren: informatie-, interpretatie en actieonzekerheden. De verhalen van ambtenaren laten zien dat zij vooral moeite hebben met de onzekerheid die onderdeel is van interacties met ondernemers, en dat zij moeten improviseren om situaties onder controle te houden of weer onder controle te krijgen. Deze improvisatie gebeurt veelal onder tijdsdruk en gaat vaak gepaard met twijfel over of er wel goed gehandeld is. Deze actieonzekerheden zijn onvoorspelbaar en vormen een controleprobleem voor contactambtenaren, en staan dus in schril contrast met de bureaucratische principes die gericht zijn op reductie van onzekerheid.

Dit proefschrift heeft bovendien laten zien dat sociale dynamieken die steeds vaker onderdeel zijn van besluitvorming over burgers, een bron van onzekerheid vormen. Onderhandelingen met burgers, maar ook besluitvorming in professionele teams, brengen sociale dynamieken mee die een doorslaggevende rol kunnen spelen in besluitvorming. Hoewel sommige respondenten

Dutch summary

doelbewust gebruik maken van hun collega's om besluitvorming minder subjectief te maken, zijn er ook impliciete sociale dynamieken die een impact hebben op besluitvorming over ondernemers. De flexibiliteit die het inzetten op onderhandelingen met burgers en besluitvorming in teams – en dus sociale dynamieken – lijkt te bieden, is tegelijkertijd een bron van onzekerheid in bureaucratische organisaties

Wanneer we kijken naar de manieren waarop ambtenaren omgaan met onzekerheid, kan er geconcludeerd worden dat zij niet alleen *on-the-spot* improviseren, en leunen op interpretaties van collega's als *back-up*, maar ook zoeken naar *bewijzen* om hun interpretaties te onderbouwen. Om een beeld te krijgen van ondernemers, kijken ambtenaren niet alleen naar beschikbare informatie, maar interpreteren ze ook signalen. Het interpreteren van signalen geeft ambtenaren inzicht in niet direct waarneembare eigenschappen, zoals de competentie of welwillendheid van ondernemers. Dit proefschrift heeft laten zien dat belastingambtenaren niet alleen kijken naar kenmerken van de boekhouding van ondernemers om hun betrouwbaarheid vast te stellen, maar ook naar meer particularistische aspecten, zoals gedrag en houding, en achtergrondkenmerken zoals sociale klasse, opleidingsniveau en etniciteit. Dit onderzoek heeft ook laten zien dat deze achtergrondkenmerken kunnen fungeren als frame dat de interpretatie van andere signalen kleurt.

About the author

Nadine Raaphorst

Nadine Raaphorst (1987) studied Sociology at the Erasmus University Rotterdam. In the third year of her bachelor, she studied abroad for one semester at the University Institute of Lisbon (ISCTE) in Portugal. After finishing her bachelor in 2008, she successfully completed the two-year research master Sociology of Culture, Media and the Arts at the same university. Within her master's thesis project she studied how Dutch general practitioners who have a holistic worldview combine conventional with alternative medicine. During her studies she worked as a teaching assistant at the Sociology department of the Erasmus University Rotterdam, and as a research assistant for the Netherlands Institute for the Study of Crime and Law Enforcement. After graduation she worked as a researcher within the field of healthcare at the University of Applied Sciences in Leiden.

In 2013, Nadine started as a Ph.D. candidate at the department of Public Administration of the Erasmus University in Rotterdam. Within her Ph.D. project she focused on frontline bureaucrats' uncertainty experiences and evaluations of citizen-clients. She used a sociological approach to study bureaucratic uncertainty and uncertainty reduction from the perspective of frontline officials. Her main research interests are the study of social context and stereotyping in bureaucracy, and the implications for decision making about citizens. During her Ph.D. Nadine also collaborated with different scholars on various themes, such as organizational socialization, the social context of street-level bureaucracy, trust within the public sector, the role of signals in frontline evaluations and ICT in the public sector. She has published her work in international peer-reviewed journals in public administration and sociology and in edited books.

Nadine was chair of the PhD platform at the department of Sociology and Public Administration at the Erasmus University from 2015 to 2016. She has tutored different work groups at the bachelor level at the same department, and has been a teaching assistant of the course Introduction to Qualitative Interpretive Methods at the ECPR Winter School in Methods and Techniques, 2016. From May to July in that year she has visited the department of Political Science at the Aarhus University, to present and discuss her work with various scholars. She has participated to several international conferences such as the European Group of Public Administration, the International Conference on Public Policy, the International Research Society for Public Management, the Public Management Research Conference and the International Meeting on Law and Society. In 2014, one of her papers got awarded the EGPA PhD Symposium Best Paper Award. She has also acted as reviewer for various peer-reviewed journals.

As of June 2017 Nadine works as assistant professor at the Faculty of Governance and Global Affairs, Leiden University.

Public organizations have traditionally been geared to reducing uncertainty by means of standardization and hierarchical control. In recent decades, managerial reforms and digitalization have made it possible to put public officials' work under even closer scrutiny and control. At the same time, however, frontline discretion is seen as essential in today's horizontal forms of service provision and law enforcement. Notions such as trust and collaboration are not predefined, but left open for frontline officials' professional expertise. This doctoral dissertation examines how frontline officials who work in a bureaucratic context and have been granted considerable discretion, experience uncertainties and attempt to reduce these uncertainties. By studying tax officials' and labor inspectors' work from a bottom-up perspective, this dissertation finds that frontline decision making is in large part dependent on indeterminate social interactions and affected by different uses of social typologies. This thesis concludes that a sociological approach to the study of frontline decision making is warranted.