

Lisa Wallander and Anders Molander

Disentangling Professional Discretion: A Conceptual and Methodological Approach

Abstract: With the aim of furthering the investigation of professional discretion, this article builds on a combination of a conceptual framework for understanding discretion and an advanced method for collecting data on human judgments. Discretion is described as consisting of two dimensions—a structural dimension (discretionary space) and an epistemic dimension (discretionary reasoning). Discretionary reasoning is defined as the cognitive activity that may take place within the discretionary space of professional judgment, and it is illustrated by means of Toulmin's model of argumentation. The factorial survey, a quasi-experimental vignette approach, is proposed and illustrated as a method with substantial potential for studying agreement and disagreement in discretionary reasoning. While the combined framework presented in this article could form the basis for case studies and/or comparative studies of discretionary reasoning across professions and contexts, the results of such studies could be used for improving practice within a specific professional field.

Keywords: discretion, discretionary reasoning, professional judgment, factorial survey, vignette, scenario

Lisa Wallander
Centre for
Profession Studies
& Department of
Social Work,
Malmö University

Anders Molander
Centre for the Study
of Professions.
Oslo and Akershus
University College
of Applied Sciences

Contact:

Lisa Wallander
Centre for
Profession Studies
& Department of
Social Work
Malmö University
205 06 Malmö
Sweden
lisa.wallander@mah.se

In the literature on professions, *discretion* is frequently portrayed as lying at the heart of professional work (e.g. Freidson, 2001; Miller, 2010, ch. 6). As expressed by Lipsky, “for some analysts[,] the defining characteristic of professionalism is simply the discretion to make decisions about clients” (Lipsky, 1980, p. 215). This view of the significance of discretion is based upon the assumption that discretion is an unavoidable aspect of the application of general knowledge, embedded in “if-then”-rules, to particular cases. All professions engage in the application of such knowledge in one form or another, and what is more they are authorized to do so. However, when general rules do not determine unambiguous conclusions about what ought to be done in particular cases, there is a space for discretion, or a “space of autonomy”, in professional judgment and decision-making (Galligan, 1986, p. 8). In the words of Ronald Dworkin, discretion is like the “hole in a doughnut,” where the circle of the doughnut comprises the “belt of restriction” (comprising standards set by various authorities), and where the hole in the middle may be larger or smaller (Dworkin, 1978, p. 31). This metaphor illustrates the common definition of discretion as an area where one can choose between permitted alternatives of action on the basis of one’s own judgments (e.g. Davis, 1969, p. 9; Barak, 1989, p. 7). In the literature on professions, discretion is in this sense frequently portrayed as something positive—as a precondition for the appropriate individualised treatment of individual cases (e.g. Handler, 1986) and as a “reflective praxis” (e.g. Schön, 1983). However, despite the fact that discretion is seen as

Received:
21 February
2014

Accepted:
3 April 2014

an unavoidable element of professional practice, it has not been without its critics (Goodin, 1988; Rothstein, 1998). It has been claimed that extensive use of discretion in the welfare state can both threaten the principles of the rule of law (such as predictability, legality and equal treatment) and undermine democratic control over the street-level implementation of laws and policies (Molander et al., 2012). In addition, scholars of cognitive psychology have shown that linear statistical models can produce better predictions than clinical (or case-based) assessments (Meehl, 1954), and that the use of heuristics may give rise to systematic errors or cognitive bias (Tversky & Kahneman, 1974; for a summary of this research tradition, see Kahneman, 2011; for an alternative approach, see for example Gigerenzer & Brighton, 2009).

In this article, we are not concerned with normative questions surrounding the use of discretion in professional work, but rather with the theoretical understanding and empirical investigation of discretion. A purely structural approach to discretion (as a space for judging, deciding and acting) says nothing about what it means to practice or exercise discretion. As a result, the hole in the doughnut might be likened to the invisible contents of a “black box”. We will be attempting to open this black box in two ways. First, we will develop a conceptual framework based on a distinction between *discretionary space* and *discretionary reasoning* (Molander & Grimen, 2010). The discretionary space, which may be viewed as the structural dimension of discretion, will not be analysed further in this article. Instead the focus will be directed at the cognitive activity carried out by an agent when he or she is making judgments and decisions under conditions of indeterminacy, i.e. discretionary reasoning. Second, we will propose the *factorial survey* [FS], a quasi-experimental vignette approach (e.g. Rossi & Nock, 1982; Jasso, 2006; Wallander, 2009), as a method with substantial potential for specifying the empirical contents of the judgments that professionals make in discretionary spaces. The factorial survey, which was originally introduced as a method for studying the shared and varying principles of human judgments (Rossi & Nock, 1982), has of late increasingly been employed in the study of professional judgments (e.g. Ludwick et al., 2004; Taylor, 2006; Wallander, 2012). Existing empirical studies include analyses of judgments made, for example, by court judges (Hagan, Ferrales, & Jasso, 2008), teachers (Webster et al., 2005), nurses (Rattray et al., 2011), physicians (Mion et al., 2010), police officers (Son, Davis, & Rome, 1998) and social workers (Wallander & Blomqvist, 2008; Samuelsson & Wallander, 2013). Although most factorial surveys conducted in the field of professional judgments make use of the concept of “professional judgment”, they are not based on a common understanding of which particular elements of such judgments can be studied using this approach. The conceptual framework that will be outlined below is proposed as having the potential to provide a common theoretical basis for all FS studies on professional judgments, irrespective of which profession is being studied.

To date, no single factorial survey study exists in which professional judgments have been dissected in all of the various ways that will be proposed below. Therefore, we will employ a design constructed for the study of social care professionals’ judgments of elder abuse (Killick & Taylor, 2012) as an example, and will draw on fictive results which have been devised for the purposes of illustration. Throughout the article, the term “client” will be used in its standard sociological meaning denoting the role that is complementary to the role of the professional.

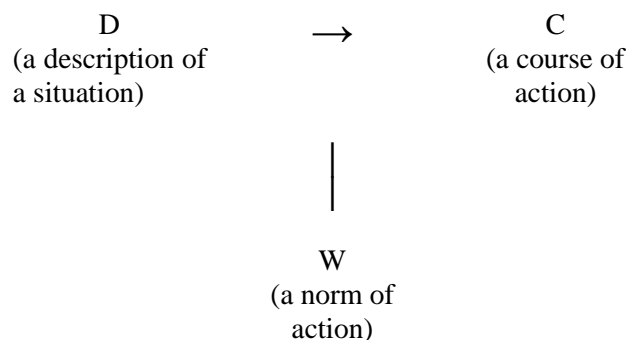
Conceptual framework – discretion as a way of reasoning

As has been mentioned above, we will not be analysing the spatial or structural dimension of discretion. We should note, however, that this dimension is inextricably linked with discretion in an epistemic sense. The entrustment of discretionary power to professionals, i.e. their being assigned a *space* for making decisions in accordance with their own judgment, is based on the assumption that discretionary judgments and decisions are not mere whimsies but are justifiable, and that the practitioners involved are capable of making *reasoned judgments and decisions*. What we expect from these professionals is that they act in accordance with their *best* judgment, which means that what they do is supported by *good* arguments. Hence, this epistemic dimension of discretion—discretion as reasoning—is fundamental from a normative point of view.¹

To reason means to attempt to find justifiable answers to questions. Professionals are concerned with practical problems, i.e. questions about what ought to be done (Gauthier, 1963). Following Toulmin's (1958) general model of argumentation, practical reasoning may be described as consisting of three components: (1) the *claim* or *conclusion* (C) refers to a course of action (or inaction), (2) the *data* (D) are a description of a situation, and (3) the *warrant* (W) has the form of a norm of action which licenses the step from the description of a situation to a conclusion about what to do. The structure of such an argument is shown in Figure 1.

Figure 1

The components of a practical argument



The *norm of action* states how we should behave or what we ought to do in situations of a specific kind. As a warrant, its role is to justify the step from a description of a situation to a practical claim about what to do. Some norms are *deontic* and specify duties. They say what one is obliged to do under certain conditions. But the step from a description of a situation to an action can also be justified by

¹The concept of discretion is sometimes used interchangeably with that of autonomy (cf. the quotation from Galligan above). On the basis of the distinction between discretionary space and discretionary reasoning, one has to differentiate between two meanings of autonomy: opportunities for judgment vs. judgmental capacity. Autonomy in the first sense becomes stronger the larger the discretionary space, and vice versa. However, since the discretionary space of professionals is constituted by acts of entrustment, there is a demand for accountability. Professionals are expected to follow their *best judgments*. The concept of autonomy that is in play here does not primarily emphasize non-interference and the permission to make use of discretion, but rather the *capacity or ability to make good judgments*. Both what constitutes a good judgment and whether a judgment is good are often controversial issues. The crucial point, however, is that the decisions and actions of an individual who lacks the capacity to make good judgments are *heteronomous*. This means that the individual's decisions and actions are governed by other persons' judgments or by tradition, not by the individual's own judgment.

norms in which the antecedent refers to an end and the consequent states an action which can realize the end. In such a case, the norms express means-ends connections and can be called *instrumental* norms or hypothetical imperatives. Still another kind of norm demands that we should realize certain ends and find adequate means to do so. Such *teleological* norms have an open structure and leave a large space for the actor's interpretations and choices. They define the weakest constraints on practical reasoning (for a typology of norms, see Schnädelbach, 1992, building on von Wright, 1963). Without norms prescribing actions one cannot make inferences about what ought to be done on the basis of a description of a situation. However, in the form of warrants, the norms can determine the conclusion to a stronger or weaker degree. Strong warrants make decisions easier—although we may nonetheless be forced to make reservations and specify additional conditions under which the conclusion is valid.

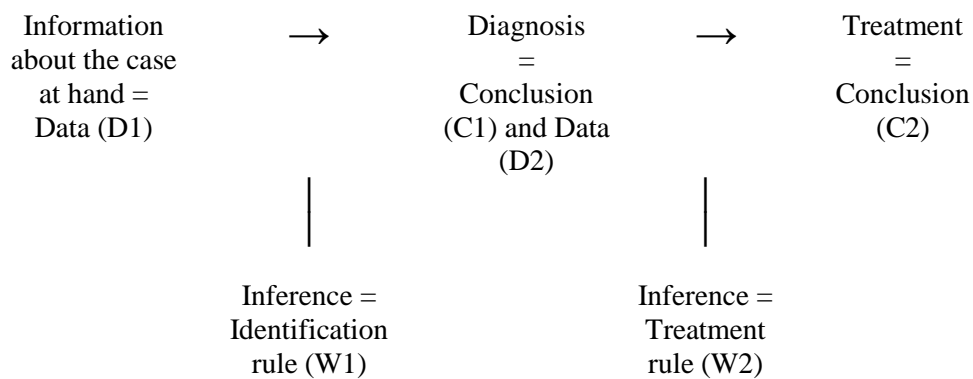
A description of a situation that constitutes the data in a practical argument is, in its turn, a conclusion in an argument about the character of the situation at hand. It identifies a situation of a certain kind. What licenses the inference from a set of data about a situation to a certain description of a situation is a *rule of identification*. If certain features occur, then there is a situation of type S. While norms of action bridge the gap between descriptions of situations and conclusions about what to do, rules of identification bridge the gap between data and conclusions about what is the case (descriptions of situations). Both types of warrants may be disputed and require justification. Justifying a warrant involves showing both that it is *valid* and that it is *applicable* in the case at issue.

Discretionary reasoning exhibits these characteristic features of practical arguments. Practical arguments may be described as forming a continuum, where the most important variable is the force of the warrants, and *discretion encompasses that part of the continuum where the warrants are weakest*. The force of a strong warrant approaches the force of a rule of deduction in logic: If the premises are true, the conclusion must be true. If all humans are mortal and if Lisa and Anders are humans, then Lisa and Anders are mortal. The conclusion is entailed by the premises. But in no domain of practical reasoning are there warrants with this kind of force. A weak warrant only mentions issues that ought to be considered in the process of reasoning. And it can be completely unspecific with regard to how these issues should be understood, considered and weighted. However weak these warrants may be, they are nonetheless in force. If there were no warrants governing reasoning, there would be no discretion, but merely “free fantasy” (Molander & Grimen, 2010, p. 174).

Let us now apply in more detail this general sketch of the structure of practical arguments to professional practice. Abbott identifies a triad of basic acts that together constitute the “essential cultural logic of professional practice” (Abbott, 1988, p. 40). The terminology that he uses in describing these acts—*diagnosis, inference, treatment*—is borrowed from medicine, but *mutatis mutandis* the terms can be used to describe professional practice in *any* given occupation.² In Figure 2, the model of practical reasoning (cf. Figure 1) is applied to the chain of diagnosis, inference and treatment.

² The term diagnosis involves forming a professional opinion about the current status of a client, while treatment is equivalent to prescription, that is, to suggesting interventions that are deemed to be the most suitable for the particular diagnosis identified (Abbott, 1988).

Figure 2
The components of discretionary reasoning in professional work



In line with the model presented in Figure 2, in order to classify (and thereby make a claim about the diagnosis/description of a particular situation/problem), professional practitioners need to combine certain forms of information about the case at hand (Data 1) with one or more “if, then”-rules warranting the conclusion about the nature of the case at issue. We have used the term *identification rules* to denote the warrants or rules of inference that are used to make claims such as these. Subsequently, in coming to a conclusion (and making a claim) about what—if any—action should be taken (in order to solve the problem(s) involved), the practitioners must combine the description of the situation, i.e. the “diagnosis” (Data 2), with one or more action norms that specifies which treatment is the most appropriate given the circumstances. These norms may be called *treatment rules*. When professionals, for example GPs and social workers, act as gatekeepers in the welfare state, they use clinical inference rules within the context of legal rules stating which individuals are eligible for certain benefits or services and what they are entitled to, if they are eligible (Molander et al., 2012). Treatment rules of the latter, legal kind correspond to the welfare rights of the citizens.³

In line with viewing practical arguments as a continuum of stronger to weaker warrants, professional reasoning may also be described as more or less discretionary. The strength of a particular warrant is dependent upon the preciseness of its components, i.e. of the antecedent (“if”) and its consequent (“then”). The antecedent may be precise, but the consequent vague; the antecedent may be vague, but the consequent precise; and both the antecedent and the consequent may be vague. In cases such as these, discretion must be used to fill the gaps (Molander & Grimen, 2010).

In Toulmin’s full model of the argument (1958), the three basic elements described above (data, warrant, conclusion) are accompanied by a further three components. Two of these—the *rebuttal* and the *qualifier*—deal with the specification of the conclusions. While conditions of rebuttal comprise various circumstances in which a particular rule is not (or is less) applicable, the qualifier indicates the overall strength of the rule for coming to a particular conclusion. Finally, one or several *backings* may be denoted, the aim of which is to justify a certain rule. In this article, we will consider all of the various elements that are potentially inherent in discretionary reasoning. In addition, we will argue that the factorial survey provides opportunities for specifying the *empirical contents* of these elements. In effect, this

³ They have the following structure (Goodin, 1988, p. 186): If some individual *I*, who satisfies certain background conditions *B*, displays characteristics *K* in circumstances *C*, then an individual *O*, who occupies official position *P*, should do *T* to or for individual *I*.

means that while the use of the FS does neither involve tracing the cognitive processes inherent in professional judgment and decision-making, nor makes it possible to identify the nature of the elements of the argument (e.g. whether a particular treatment rule is based on a clinical inference or on a legal rule), it nonetheless allows us to study the contents of the elements that make up the practitioners' arguments in relation to a particular professional judgment. Moreover, while the concept of discretion has traditionally been used to describe a space for potential variation in judgments (cf. discretionary space), the factorial survey allows us to investigate both *agreement* and *disagreement* in discretionary judgments. In this context, agreement relates to the elements of a particular argument that are *common* to practitioners when coming to conclusions about particular cases. Disagreement, on the other hand, signifies the *variation* in these elements between individual practitioners and between individuals working in different contexts or organizations.

Methodological framework – the factorial survey

The central components of factorial surveys are the vignettes that are judged by the respondents. For respondents in a factorial survey study, vignettes constitute fictive descriptions of people or social situations. For the researchers, the vignettes primarily represent different combinations of *levels* (values) of various *dimensions* (variables), which are included on account of their expected relevance as determinants of the judgment of interest (Rossi & Nock, 1982). Figure 3 presents an example of a vignette from a study of social care professionals' (e.g. social workers', nurses') recognition and reporting of elder abuse. With the overall aim of contributing to a better understanding of professional conceptualisations and decision-making in relation to the abuse of the elderly, this study analysed multiple judgments made by 190 social care professionals in Northern Ireland (Killick & Taylor, 2012). As may be noted, the vignette in the example is succeeded by two rating scales, the first of which relates to the respondents' definition of the situation as potentially involving abuse (diagnosis), while the second relates to the respondents' estimations of their likelihood of referring the case for investigation (treatment).

Figure 3

An example of a factorial survey vignette (Killick & Taylor, 2012)

Your client is a 74 year old female who has had a minor stroke. She can sometimes be confused and is demanding. She is looked after by a daughter who finds the role stressful and has unrealistic expectations. The daughter admits that she punished her with a slap on two occasions. The client consents to an investigation. The daughter will give up the caring role if an investigation is initiated. There are no available day care or respite places.

To what extent do you perceive this to be abuse?

Not Abuse 0 1 2 3 4 5 6 7 8 9 Abuse

How likely would you be to refer this case for investigation?

Not Likely 0 1 2 3 4 5 6 7 8 9 Very Likely

The vignette design of the study consists of twelve dimensions, including (1) client age, (2) client sex, (3) client condition, (4) client capacity, (5) client behaviour, (6) carer stress, (7) carer factor, (8) type of act, (9) frequency of act, (10) client wishes, (11) carer outcome and, (12) resources. As may be noticed, these dimensions include both dimensions describing the characteristics of the people involved in a particular situation, and dimensions describing the situation and its context. Each dimension contains a specific number of levels (for an overview, see Table A1 in Appendix). For example, “client condition” comprises five levels (eczema; diabetes; severe arthritis; had a minor stroke; had a major stroke), whereas “type of act” comprises four levels (roughly handled him/her; shook him/her by the shoulders; punished him/her with a slap; hit him/her in the face with a fist). In FS terminology, the compilation of all combinations of dimension levels—making up the maximum number of unique vignettes—is referred to as the *vignette universe*. The vignette universe of the design employed in this study consists of the product of the number of levels attached to the twelve dimensions, that is $4 \times 2 \times 5 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 = 10,485,760$ different vignettes.

In a factorial survey, the respondents do not individually judge all of the objects comprising the vignette universe, but samples of vignettes, which are drawn by means of a random or a quota sampling design (Dülmer, 2007). In the study used here as an example, each respondent judged a (uniquely drawn) random sample of 12 vignettes, producing a total of 2,261 judgments at the aggregate level. As a consequence of the sampling procedures employed in factorial surveys, the correlations between the dimensions of the vignettes included in the sample are as a rule approximate to zero. This experimental characteristic makes it possible to disentangle the unique effects on judgments of dimensions that may in reality be highly correlated, and it thereby produces study results with a high level of internal validity. In addition, the fact that the FS makes it possible to simultaneously examine and control for a large number of variables and values (in contrast to the more well-known factorial experiment), suggests that the results will also have relatively high levels of external validity (Rossi & Nock, 1982).⁴

Historically, the examination of the data produced in factorial surveys has most commonly been carried out by means of ordinary multiple regression analysis. More recently, however, the multilevel extension of regression analysis has come to be viewed as the new standard for factorial survey analyses (Wallander, 2009, 2012). This is due to the fact that the respondents in an FS study generally judge multiple vignettes, which means that the dataset as a result becomes hierarchical by design.⁵ In addition, studies of professional judgments are often based on hypotheses associated with assumptions that practitioners’ judgments might not only be affected by characteristics related to the clients and the practitioners themselves, but also by factors related to their work

⁴ Although the factorial survey enjoys the advantages of both the experiment and the social survey (Rossi & Nock, 1982), and notwithstanding the fact that more than 100 articles using this approach have been published in sociology journals during the last three decades (Wallander, 2009), it remains unknown to most social scientists. In fact, the approach was not even included among those covered in a recently published book about population-based survey experiments (Mutz, 2011). However, in other disciplines, not least in health economics, the discrete-choice experiment—a method that is in many respects similar to the factorial survey—is steadily gaining ground (De Bekker-Grob, Ryan & Gerard, 2012).

⁵ This means that units in the dataset are clustered, or nested, within units at a higher level (Snijders, 2004).

contexts. Some of these studies have employed clustered sampling techniques (e.g. Wallander & Blomqvist, 2008), in line with which the initial sampling frame consists of workplaces rather than of individuals. Such a strategy produces a dataset in which the respondents are clustered within their workplaces, and which therefore reflects a “natural” hierarchy. The analytical strategy presented in this article builds on the advantages afforded by multilevel analysis, including for example the possibility to correctly specify whether and if so how professional judgments are contextually structured, and it presupposes a dataset that is hierarchically structured along three levels: 1) the client/vignette, 2) the practitioner, 3) the work context (for further details about the multilevel modelling of FS data, see Hox, Kreft, & Hermkens, 1991; Wallander, 2008, 2012; for details about another strategy that is also very useful in the analysis of FS data, see Jasso, 2006).

Analysing agreement in discretionary reasoning

As was noted above, our methodological framework involves analysing both which elements of discretionary reasoning are common to practitioners (agreement) and potential variation in these elements between individual practitioners and between individuals working in different contexts (disagreement). The subsequent presentation of analyses will be structured in accordance with this distinction. Throughout the presentation, the design used in the above mentioned study of social care professionals’ judgments about elder abuse (Killick & Taylor, 2012) will be used to exemplify our arguments. However, since the analytical strategy employed in this recently published study differs from that which we propose in this article, the analyses and results presented below are not entirely true to the original, but have been modified for the purposes of illustration.

In portraying the various elements of discretionary reasoning that can be studied in a factorial survey, it is important to bear in mind that the data that are used as a point of departure for a particular claim are entirely determined by the information incorporated into the vignettes judged by the respondents. Similarly, the number of feasible conclusions is determined by the rating tasks that follow the vignettes. Taking the study by Killick and Taylor (2012) as an example, the data consist of the twelve characteristics that together describe a situation of potential elder abuse and the two people involved in it. The conclusions to be analysed are bounded by a fixed number of grades on two rating scales, which allow the respondents to judge the degree to which they perceive the situation to be abuse (diagnosis) and the degree to which they find it likely that they would refer the case for investigation (treatment).

The identification and treatment rules that are more or less common to the respondents’ judgments may be inferred from analyses of the effects of the vignette dimensions (the independent variables) on respondents’ aggregate judgments (the dependent variables). Such analyses are typically carried out by means of ordinary regression analysis. However, our methodological framework involves hierarchical data and presupposes the use of multilevel regression analysis (see above). The output from multilevel regression models is split into two parts: the fixed part, corresponding roughly to the output from standard regression analysis, and the random part, comprising the decomposition of the unexplained variance into variance components for each level of the data set. As the results associated with agreement in discretionary reasoning are displayed in the fixed part of the model, the interpretation of these results is similar to that of results from ordinary regression analysis.

Now, let us construct an example based on the elder abuse design employed by Killick and Taylor (2012). One of the dimensions of the design is “type of act”, and it is represented by the following levels: a) the carer roughly handled him/her; b) the carer shook him/her by the shoulders; c) the carer punished him/her with a slap; d) the carer hit him/her in the face with a fist. Let us consider an analysis which shows that punishment with a slap is on average more often recognized as elder abuse (on a scale between 0 and 9) than rough handling. Such a result can be expressed in the form of an identification rule such as the following: *if the situation involves punishment with a slap, it is more likely to be a case of elder abuse, by comparison with situations involving rough handling.*⁶ A similar effect of this dimension on the respondents’ judgments about reporting would translate into the following treatment rule: *if the situation involves punishment with a slap, I would be more likely to refer the case for investigation, by comparison with situations involving rough handling.* These identification and treatment rules may constitute examples of rules that are indeed consciously used by many of the respondents when making judgments about the vignettes. However, one of the benefits of using this method is that it also allows for the detection of tacit and/or “prejudiced” rules that are used in discretionary reasoning (Jasso, 1998; Wallander, 2012). Let us imagine, for instance, that the following rule was to be uncovered in the analyses: *if the situation involves a female client, it is more likely to be a case of elder abuse, by comparison with situations involving a male client.* It is a plausible assumption that the professionals who employed this rule—in accordance with which the recognition of elder abuse is dependent on the sex of the victim—would either be unaware of it (thereby using a tacit/implicit rule), or if asked about it, would be reluctant to admit to it, thereby avoiding what might be regarded as a prejudice. One of the chief advantages of the factorial survey relates to the fact that the respondents are likely to be either inattentive to, or unable to obtain a complete overview of, the experimental manipulation of the vignette dimensions and that it therefore allows the detection of results that would have been difficult to obtain using other data-collection methods (e.g. methods involving the use of fully identical and non-varying vignettes across respondents). This makes it particularly suitable for investigating the implicit employment of social and cultural stereotypes in discretionary reasoning.

As was described above, claims in discretionary reasoning are generally specified through the use of conditions of rebuttal and qualifiers. While conditions of rebuttal comprise various circumstances in which the rule is not (or is less) applicable, the qualifiers indicate the overall strength of a rule for coming to a certain conclusion. In studies using a factorial survey design, both these components may be empirically established. First, conditions of rebuttal that are to some extent shared by the respondents may be investigated by including interaction terms as predictors in the analysis.⁷ Let us consider an analysis which shows that the effect of “type of act” on professionals’ recognition of elder abuse (see above) is moderated by the “behaviour” of the client. Thus, while respondents might make a distinction in judgments between punishment

⁶ It might be argued to be more accurate to formulate a rule as follows: “if the situation involves punishment with a slap, it is likely to be a case of elder abuse”. However, because the rules identified in FS studies are based on an analysis of respondents’ weightings of levels within vignette dimensions (which are most often categorical variables), the category or categories of comparison within a specific dimension must be incorporated into the rule.

⁷ Only rebuttals associated with the other vignette dimensions may be investigated in such an analysis.

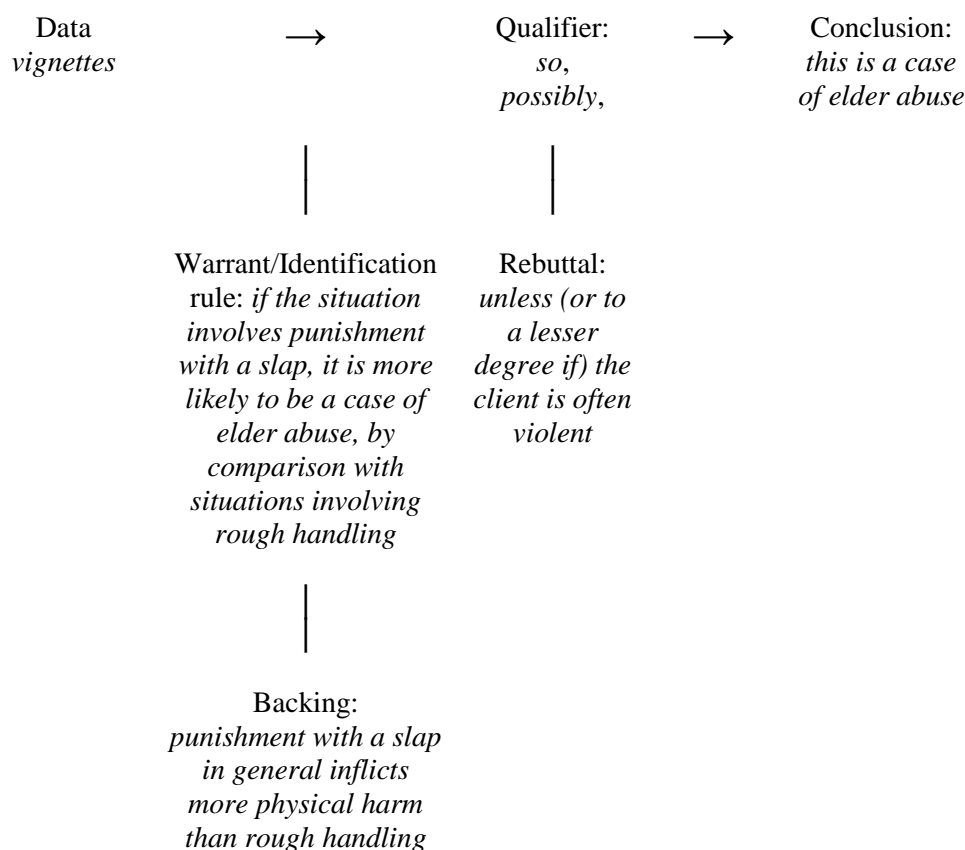
with a slap and rough handling in situations where the client is placid, demanding or aggressive, this might not be the case (or at least not to the same extent) in situations where the client is often violent. Were this to be the case, it would mean that one of the rebuttals to the above described identification rule would comprise situations in which *the client is often violent*.

The specification of the qualifier is not quite as straightforward, and it is partly subject to the discretion of the researchers involved in the study. The independent force of each rule in guiding the overall conclusions—signifying the degree to which the respondents agree on the rule—is given by the magnitude of the standardized regression coefficient that represents the effect of a particular vignette dimension (or the difference between two levels within a particular dimension). However, given that the qualifier indicates how certain one can be about a specific conclusion—given the use of a particular rule—it must also allow for potential moderations of the main effect, i.e. conditions of rebuttal. Accordingly, after considering both the main effect of a rule and potential exceptions to the rule, the researchers working on the study may decide on an adverb that correctly expresses the overall strength of the rule. Such adverbs vary from being fairly strong, such as “probably” or “presumably”, to weak, e.g. “potentially” or “possibly” (cf. Toulmin, 1958).

Notwithstanding the great flexibility of the factorial survey, it falls short of uncovering possible common backings—i.e. shared justifications—of the rules identified in the analysis. This shortcoming of the method may be dealt with in several ways. First, earlier research and theory may be used to inform the overall understanding of which justifications, or motives, may have served as a basis for practitioners’ use of rules in coming to conclusions about the vignette cases. One such potential justification for the rule used as the main example above might be that *punishment with a slap in general inflicts more physical harm than rough handling*. One way of empirically examining the common backings used in relation to rules is to supplement the factorial survey with qualitative methods, such as individual and/or group interviews for example. In such interviews, the practitioners could be asked to suggest one or more backings that may legitimise the application of certain rules to particular data.

The analysis of agreement in discretionary reasoning, as examined by the use of a factorial survey, is summarized in Figure 4.

Figure 4
*A model of discretionary reasoning based on fictive results from a factorial survey on elder abuse (research design by Killick & Taylor, 2012)**



* The formulation of the conclusion may give the impression that the respondents in the study have only been given two possible alternatives to choose from (*elder abuse* or *not elder abuse*), whereas in reality the rating task involved a ten-point-scale. However, while the reasons for choosing a continuous rating task for a study design in general involve allowing for more variation in respondents' judgments (and hence a more detailed statistical analysis), the phrase chosen for the model in Figure 4 (which corresponds to one of the two end-points of the scale) is more faithful in its wording to the typical claims made in arguments.

Analysing disagreement in discretionary reasoning

At the general level, factorial survey data allow for multiple ways of analysing variation in judgments (Byers & Zeller, 1998). However, in this article, we are primarily interested in analysing disagreement in *conclusions* as well as in the *use of rules* in coming to conclusions. Moreover, we want know whether such potential variation is mainly due to differences between individual practitioners or sub-groups of practitioners, or whether some of the disagreement in judgments may also be ascribed to the work context.

While the interpretation of agreement in discretionary reasoning is completely based on the fixed part of the multilevel model, the analysis of disagreement makes full use of the possibility afforded by multilevel analysis to analyse in detail the variance that remains unexplained in a specific model. This unexplained variance is decomposed into so-called *variance components* for each of the levels of the

design. When unexplained variance associated with the *intercept* is detected, it is possible to draw the conclusion that the average claims based on the vignette cases vary systematically, either between individual practitioners (at Level 2) or between individuals working in different contexts (at Level 3). When unexplained variance linked to the *regression slope(s) representing a particular vignette dimension* is identified, it is possible to conclude that the use of the rule(s) associated with that particular vignette dimension varies systematically, either between individual practitioners (at Level 2) or between individuals in different work contexts (at Level 3). Although the possibility of specifying the level of the unexplained variance in itself provides opportunities for many interesting interpretations, the next step of the analysis is logically to try to explain this variance, by including variables relating to the practitioners and/or to the work contexts as predictors in the analysis.

The analysis of disagreement in conclusions begins by running a so-called “empty” model (i.e. a model without predictors). Let us suppose that such a model was applied to the social care professionals’ judgments about the recognition of elder abuse (Killick & Taylor, 2012), and that the variance components associated with the intercept were significant at both Level 2 and Level 3. This would indicate the existence of individual and contextual disagreement in the propensity to make the claim that a certain situation constitutes elder abuse. By including respondent and contextual variables as predictors in the model, such disagreement might be partly or fully explained. Consider an example where respondents who are nurses on average make higher ratings on the abuse-recognition scale than those who are social workers. This would mean that the professional affiliation of the respondents would explain some of the individual disagreement in the propensity to recognize a case as elder abuse. Further, if a contextual variable, such as specific guidelines for practice (implemented at the level of the workplace), were to have an influence on respondents’ overall ratings, this would mean that some of the contextual disagreement in the propensity to recognize a case as elder abuse would be explained by contextual variation in the use of these guidelines.

The analysis of disagreement in the use of rules/warrants begins by focusing on a particular rule of interest, and by analysing the variance component that relates to the regression coefficient associated with this rule. Let us suppose, for example, that there is unexplained variance at the respondent-level of the analysis in the use of the following rule, associated with the condition of the client: *if the situation involves a client who has had a major stroke, it is more likely to be a case of elder abuse, in comparison with situations which involve a client who has eczema*. The second step of this analysis would be to include one or more variables describing the respondents as determinants of the regression slope that represents this difference in judgments (these terms are commonly referred to as *cross-level interactions*). If we once again use the “professional group” respondent variable as an example, such an analysis might show that nurses on average use the rule described above more often than social workers. Taken together, these analyses show ways of identifying and explaining individual disagreement in the use of rules for recognizing elder abuse. Obviously, an identical analytical strategy would be used for identifying and explaining contextual disagreement in the use of rules (interpreting variance components at the contextual level of the analysis and including contextual predictors in cross-level interactions).

Analysing the relationship between diagnosis and treatment

As part of the conceptual framework outlined in this article, we linked the model of the argument to the triad of professional acts proposed by Abbott (1988). For our purposes, the act of inference was decomposed into treatment rules that are used to bridge the gap between the act of diagnosis (data) and the act of treatment (conclusion; cf. Figure 2). However, since a diagnosis in itself

is (or at least should be) a constant, it cannot be used as the only data to be considered in an analysis of conclusions about treatments. A prerequisite for the empirical specification of treatment rules is that the information that is used as data includes many *different* dimensions that may be linked to judgments about treatment. Thus, when using the factorial survey to analyse the relationship between diagnosis and treatment, it is essential to proceed on the basis of one and the same vignette design, and to treat the respondents' judgments about diagnosis as a variable that may or may not be mediating the effects of the other vignette dimensions on their judgments about treatment. This would involve a step-wise regression analysis, in which the first step estimates the effects of the vignette dimensions on the judgments about treatment, thereby identifying "provisional" common treatment rules. The second step would involve incorporating the respondents' judgments about diagnosis as a further independent variable in the analysis and analysing the results that follow.

If we take the elder abuse design as an example (Killick & Taylor, 2012), a first such analysis might show that vignette dimensions associated with client characteristics, the situation involved and client wishes (whether or not action should be taken) have an influence on the social care professionals' estimations of their likelihood of referring a case for investigation. A second analysis, in which the practitioners' judgments regarding the recognition of elder abuse was included as an additional independent variable, might show that this new variable has a strong effect on judgments about referrals, with the effects of client and situational characteristics perhaps disappearing, but with client wishes continuing to have an effect. This would mean that there is a strong relationship between the respondents' diagnosis (recognition) and treatment (referrals for investigation) of elder abuse, but that the social care professionals are also guided by additional treatment rules associated with the wishes of the clients.

As far as the judgments used as examples throughout this article are concerned—i.e. social care professionals' judgments about the recognition of elder abuse and their estimations of their likelihood of referring the case for investigation—we might right from the start expect to find a strong relationship between diagnosis and treatment. However, there are countless instances in which we would be less certain about the strength of the actual relationship between professionals' diagnosis and treatment of clients. For example, when selecting a treatment for a specific diagnosis, doctors generally have several options to choose between. This is also the case for teachers and psychologists, who make recommendations regarding the need for special education for pupils who do not satisfactorily benefit from ordinary tuition, and for social workers, whose task it is to suggest interventions for problem substance users. These are of course just three examples drawn from a wide range of professional judgments that would merit further examination.

Concluding remarks

With the aim of opening up the "black box" of discretion, we have proposed a combination of a conceptual framework, based on the distinction between discretionary space and discretionary reasoning, and a methodological framework, in the form of the factorial survey and the use of multilevel regression analysis. We have described discretionary reasoning as the cognitive activity that may take place within the discretionary space of professional judgment and decision-making. This activity has been illustrated by means of Toulmin's model of argumentation. Further, we have described ways of using the factorial survey to analyse agreement and disagreement in discretionary reasoning. In the following, we will discuss some of the implications associated with our suggestions.

First, although there are many benefits associated with linking, as we do, a conceptual analysis of discretion with a systematic approach for empirically studying the subject matter, there are also a number of associated caveats. Most importantly, at the conceptual level, we have modelled discretionary reasoning on the basis of an assumption that conclusions are accepted on the basis of justified reasons. However, in studies using the factorial survey, we cannot know whether the respondents' judgments are actually the result of reasoning, or whether they are more or less intuitive judgments. This also means that we cannot judge the quality of their reasoning, i.e. whether they are making mistakes or are subject to biases (for a summary of the work in the psychology of reasoning, see Mercier & Sperber, 2011). Another caveat involves the fact that the judgments, as measured in an FS study, are made in a non-communicative setting, in which the respondents do not have to justify their judgments to others. If we want to trace the cognitive processes of the respondents, or study whether an argumentative or deliberative setting influences their discretionary reasoning (cf. Sunstein, 2006: 45 ff; Mercier & Sperber, 2011; Landemore, 2012; Sperber & Mercier, 2012), we must use other data-collection methods (e.g. in combination with the factorial survey), such as the "think-aloud" or "talk-aloud" protocols, for example (Ericsson & Simon, 1993), or various experimental methods. In view of this, we should bear in mind that the use of the factorial survey involves treating the judgment that is being studied as one of several "frozen moments" (Wallander, 2008, p. 60) from a chain of discretionary judgments in professional practice, and that it allows us to model the empirical contents of the elements that together make up the structure of this particular judgment.

In this article, we have made a distinction between agreement and disagreement in discretionary reasoning. Agreement denotes factors that are taken into account and used in a similar way by many practitioners when arriving at conclusions about diagnosis and treatment. Of course, such common factors—as identified in factorial survey analyses—may reflect "false" assumptions and do not in themselves legitimize diagnoses or treatments, but they could be utilized in a number of different ways. First, they could fruitfully be employed as validated hypotheses to be further tested in research. For example, common rules denoting the suitability of different treatments for different clients could be used as a means of structuring studies of client-treatment matching effects. Second, they could be included in the material used in panel discussions involving recognized experts on a particular subject or in discussions in collegial forums. Such discussions might generate suggestions regarding a range of common justifications for the empirically identified identification and treatment rules (and potential rebuttals). In combination, results from a factorial survey and expert panel discussions could form the basis for improving practice in a specific professional field.

Every now and then, when some new evidence emerges suggesting that professionals disagree in their judgments, heated discussions commence in various social and political arenas, often involving a questioning of the "professionalism" of the individuals concerned and/or the occupational group to which these people belong. However, it is a feature of discretionary reasoning that even individuals who reason as conscientiously as possible may arrive at different conclusions about diagnosis and treatment (Molander & Grimen, 2010). Hence, care should be taken in order to thoroughly investigate the actual causes of such variation. The use of multilevel modelling within the factorial survey research design, as described in this article, makes it possible to systematically describe and explain potential disagreement in the various elements of discretionary reasoning. Existing studies employing this combination of methods have shown that judgments (i.e. conclusions) about the treatment of a case may indeed vary in line with characteristics associated with the practitioners themselves, such as their treatment approaches or "ideologies" for example (Wallander & Blomqvist, 2005, 2008), task specialization (Degenholtz et al., 1999) and work role (Wallander & Blomqvist, 2008), and also in line with fac-

tors associated with the work context, such as the actual supply of treatments, local guidelines (Wallander & Blomqvist, 2008), workplace specialization and the level of experience of carrying out particular types of investigation (Wallander & Blomqvist, 2005). Moreover, the capacity of multilevel modelling to specify the level of the unexplained variance allows the researchers working with a study to draw tentative conclusions about the potential effects on discretionary reasoning of contextual factors that may be difficult to measure and include in statistical analyses, such as workplace norms and informal routines for example.

In conclusion, we would like to suggest that the above-described conceptual and methodological approach to the study of professionals' discretionary judgments might well form the basis for numerous studies examining issues of topical interest for scholars of professional groups. These might, for example, include comparisons of discretionary reasoning across professions, professional careers, time and space (e.g. countries and/or organizations), and might also investigate questions relating to the presence of social and cultural stereotypes in the practice of discretion, the influence of the profession vs. the workplace (i.e. the organization) on the claims that professionals make, similarities and differences in judgments between novices and experts, and the development of and competition over professional jurisdictions (cf. Abbott, 1988).

Funding

This work was supported by Malmö University and Oslo and Akershus University College of Applied Sciences.

Appendix

Table A1

Factorial survey design in Killick & Taylor (2012)

Dimension	Dimension level
Characteristics of the client	
Age	65 year old 74 year old 86 year old 93 year old
Sex	Male Female
Condition	Eczema Diabetes Severe arthritis Had a minor stroke Had a major stroke
Capacity	Is very confused Can sometimes be confused Shows no confusion NULL*
Behaviour	Is placid Is demanding Is aggressive Is often violent

Characteristics of the carer	
Carer stress	Copes well Finds the role stressful Is under immense stress NULL*
Carer factor	Abuses alcohol Has a mental illness Is financially dependent Has unrealistic expectations
Characteristics describing the situation	
Type of act	Roughly handled him/her Shook him/her by the shoulders Punished him/her with a slap Hit him/her in the face with a fist
Frequency of act	One occasion Two occasions Three occasions Many occasions
Other characteristics	
Client wishes	The client wishes action to be taken The client consents to an investigation The client does not wish action to be taken NULL*
Carer outcome	The daughter will be devastated if an investigation is initiated The daughter will give up the caring role if an investigation is initiated The daughter will make a formal complaint if an investigation is initiated NULL*
Resources	A range of support services are currently available There is a six month waiting list for services There are no available day care or respite places NULL*

*A null category allows the dimension in question to be removed from a specified proportion of the vignettes in the sample.

References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. Chicago: The University of Chicago Press.
- Barak, A. (1989). *Judicial discretion*. New Haven: Yale University Press
- Byers, B. & Zeller, R. A. (1998). Measuring subgroup variation in social judgment research: A factorial survey approach. *Social Science Research*, 27(1), 73–84. <http://dx.doi.org/10.1006/ssre.1997.0613>
- Davis, K. C. (1969). *Discretionary justice: A preliminary inquiry*. Baton Rouge: Louisiana State University Press.
- De Bekker-Grob, E. W., Ryan, M. & Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics*, 21(2), 145-172. <http://dx.doi.org/10.1002/hec.1697>

- Degenholtz, H. B., Kane, R. A., Kane, R. L., & Finch, M.D. (1999). Long-term care case managers' out-of-home placement decisions: An application of hierarchical logistic regression. *Research on Aging*, 21(2), 240–274.
<http://dx.doi.org/10.1177/0164027599212005>
- Dülmer, H. (2007). Experimental plans in factorial surveys: Random or quota design? *Sociological Methods & Research*, 35(3), 382–409.
<http://dx.doi.org/10.1177/0049124106292367>
- Dworkin, R. (1978). *Taking rights seriously*. Cambridge (Mass.): Harvard University Press.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised edition). Cambridge (Mass.): MIT Press.
- Freidson, E. (2001). *Professionalism: The third logic. On the practice of knowledge*. Chicago: University of Chicago Press.
- Galligan, D. J. (1986). *Discretionary powers: A legal study of official discretion*. Oxford: Clarendon Press.
- Gauthier, D. (1963). *Practical reasoning*. Oxford: Clarendon Press.
- Gigerenzer, G. & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in Cognitive Science*, 1, 107–143.
<http://dx.doi.org/10.1111/j.1756-8765.2008.01006.x>
- Goodin, R. (1988). *Reasons for welfare: The political theory of the welfare state*. Princeton: Princeton University Press.
- Hagan, J., Ferrales, G., & Jasso, G. (2008). How law rules: Torture, terror, and the normative judgements of Iraqi judges. *Law & Society Review*, 42(3), 605–644.
<http://dx.doi.org/10.1111/j.1540-5893.2008.00353.x>
- Handler, J. F. (1986). *Conditions of discretion: Autonomy, community, bureaucracy*. New York: Russel Sage Foundation.
- Hox, J. J., Kreft, I. G. G., & Hermkens, P. L. J. (1991). The analysis of factorial surveys. *Sociological Methods & Research*, 19(4), 493–510.
<http://dx.doi.org/10.1177/0049124191019004003>
- Jasso, G. (1998). Exploring the justice of punishments: Framing, expressiveness, and the just prison sentence. *Social Justice Research*, 11(4), 397–422.
<http://dx.doi.org/10.1023/A:1022171207173>
- Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research*, 34(3), 334–423.
<http://dx.doi.org/10.1177/0049124105283121>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Killick, C., & Taylor, B. J. (2012). Judgements of social care professionals on elder abuse referrals: A factorial survey. *British Journal of Social Work*, 42(5), 814–832. <http://dx.doi.org/10.1093/bjsw/bcr109>
- Landemore, H. (2012). Democratic reason: The mechanism of collective intelligence in politics. In Landemore, H. & Elster J. (eds.) *Collective wisdom: Principles and mechanisms* (pp. 251–289). Cambridge: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511846427.012>
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russel Sage Foundation.
- Ludwick, R., Wright, M. E., Zeller, R. A., Dowding, D.D., Lauder, W., Winchell, J. (2004). An improved methodology for advancing nursing research: Factorial surveys. *Advances in Nursing Science*, 27(3), 224–238.
<http://dx.doi.org/10.1097/00012272-200407000-00007>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
<http://dx.doi.org/10.1037/11281-000>

- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74.
<http://dx.doi.org/10.1017/S0140525X10000968>
- Miller, S. (2010). *The moral foundations of social institutions: A philosophical study*. Cambridge: Cambridge University Press.
- Mion, L. C., Sandhu, S. K., Khan, R. H., Ludwick, R., Claridge, J.A., Pile, J., ... Winchell, J. (2010). Effect of situational and clinical variables on the likelihood of physicians ordering physical restraints. *Journal of American Geriatrics Society*, 58(7), 1279-1288. <http://dx.doi.org/10.1111/j.1532-5415.2010.02952.x>
- Molander, A. & Grimen, H. (2010). Understanding professional discretion. In Svensson L. G. & Evetts J. (eds.) *Sociology of professions: Continental and Anglo-Saxon traditions* (pp. 167-187). Göteborg: Daidalos.
- Molander, A., Grimen, H. & Eriksen, E. O. (2012). Professional discretion and accountability in the welfare state. *Journal of Applied Philosophy*, 29(3), 214-230. <http://dx.doi.org/10.1111/j.1468-5930.2012.00564.x>
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton: Princeton University Press.
- Rattray, J. E., Lauder, W., Ludwick, R., Johnstone, C, Zeller, R., Winchell, J. ... Smith, A. (2011). Indicators of acute deterioration in adult patients nursed in acute wards: A factorial survey. *Journal of Clinical Nursing*, 20(5-6), 723-732. <http://dx.doi.org/10.1111/j.1365-2702.2010.03567.x>
- Rossi, P. H. & Nock, S. L. (1982). *Measuring social judgments: The factorial survey approach*. Beverly Hills: Sage.
- Rothstein, B. (1998). *Just institutions matter: The moral and political logic of the universal welfare state*. Cambridge: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511598449>
- Samuelsson, E. & Wallander, L. (2013). Disentangling practitioners' perceptions of substance use severity: A factorial survey. *Addiction Research & Theory*, Early Online. <http://dx.doi.org/10.3109/16066359.2013.856887>
- Schnädelbach, H. (1992). Rationalität und Normativität. In Schnädelbach H. (ed.) *Zur Rehabilitierung des animal rationale* (pp. 79-103). Frankfurt am Main: Suhrkamp.
- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Snijders, T. A. B. (2004). Multilevel analysis. In Lewis-Beck M. S., Bryman A. & Futing Liao T. (eds.) *The SAGE encyclopedia of social science research methods* (pp. 673-677). Thousand Oaks: Sage.
<http://dx.doi.org/10.4135/9781412950589.n590>
- Son, I. S., Davis, M. S. & Rome, D. M. (1998). Race and its effect on police officers' perceptions of misconduct. *Journal of Criminal Justice*, 26(1), 21-28.
[http://dx.doi.org/10.1016/S0047-2352\(97\)00051-2](http://dx.doi.org/10.1016/S0047-2352(97)00051-2)
- Sperber, D. & Mercier, H. (2012). Reasoning as social competence. In Landemore H. & Elster J. (eds.), *Collective wisdom: Principles and mechanisms* (pp. 368-392). Cambridge: Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511846427.016>
- Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford: Oxford University Press.
- Taylor, B. J. (2006). Factorial surveys: Using vignettes to study professional judgement. *British Journal of Social Work*, 36(7), 1187-1207.
<http://dx.doi.org/10.1093/bjsw/bch345>
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(27), 1124-1131.
<http://dx.doi.org/10.1126/science.185.4157.1124>

- von Wright, G. H. (1963). *Norm and action*. London: Routledge.
- Wallander, L. (2008). *Modelling professional judgements: An application of the factorial survey to the field of social work*. Stockholm: Stockholm University.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, 38(3), 505–520.
<http://dx.doi.org/10.1016/j.ssresearch.2009.03.004>
- Wallander, L. (2012). Measuring social workers' judgements: Why and how to use the factorial survey approach in the study of professional judgements. *Journal of Social Work*, 12(4), 364-384. <http://dx.doi.org/10.1177/1468017310387463>
- Wallander, L. & Blomqvist, J. (2005). Who “needs” compulsory care? A factorial survey of Swedish social workers' assessments of clients in relation to the “Care of abusers (special provisions) Act”. *Nordic Studies on Alcohol and Drugs*, 22(English supplement), 63–85.
- Wallander, L. & Blomqvist, J. (2008). Modeling ideal treatment recommendations: A factorial survey of Swedish social workers' ideal recommendations of inpatient or outpatient treatment for problem substance users. *Journal of Social Service Research*, 35(1), 47–64. <http://dx.doi.org/10.1080/01488370802477436>
- Webster, S. W., O'Toole, R., O'Toole, A. W. & Lucal, B. (2005). Overreporting and underreporting of child abuse: Teacher's use of professional discretion. *Child Abuse & Neglect*, 29(11), 1281-1296.
<http://dx.doi.org/10.1016/j.chiabu.2004.02.007>