The Hebrew University of Jerusalem

Federmann School of Public Policy and Government

A Formal Model of Social Blame

Master Thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Public Policy

Supervised by Dr. Raanan Sulitzeanu-Kenan

Submitted by Lior Sheffer (ID: 066093675)

September 2010

Abstract

Politicians are constantly subject to instances of public blaming and their behavior is constantly affected by it. Although blame is extensively explored by psychologists and political scientists, the current body of literature lacks a unifying framework that incorporates personal-level blame attribution into a macro-level explanation of the formation of social blame – a framework that enables an analysis of the social and political conditions which alter the intensity and personal focus of blame attributions.

This work attempts to bridge this theoretical gap by asking what structural links shape the relationship between social blame-makers and blame-takers. To provide a possible answer, we present a formal model of political blame attribution that captures how individual blame attribution is transformed into a notion of aggregate blame, relying on previous psychological descriptive models of blame formation. Following this, we use the model to explore a possible connection between two central political science constructs – institutional design and social cleavages – and the way social blame is shaped. For this purpose we introduce concepts of coherence and agreement, and provide formal measures for these concepts based on our definitions of social blame. This typology then serves as a potential guideline for future research.

Contents

1. Introduction

The attribution of blame is ubiquitous in politics. Blame motivates various patterns of response by political actors, and the act of blaming is an inseparable part of citizens' engagement with the world of politics and with politicians. The formation of blame, its constructs and its effects are of interest to both psychologists and political scientists.

However, existing psychological literature tends to focus on cognitive and emotional processes that are involved in individual and interpersonal relations, rather than the broader formations that constitute social blame. In political science, different researches have attempted to identify the underlying constructs of social blame. Despite the advancement of several different conceptualizations, few works have explicitly described the dynamics of social blame formation or provided an analysis of the social or political conditions which amplify or reduce the intensity and personal focus of blame attributions.

The current body of literature therefore lacks a unifying framework that incorporates the psychological insights on how feelings of blame emerge within individuals, into a macro-level explanation of the formation of social blame. In this work we attempt to bridge this theoretical gap.

This thesis presents a formal model of social blame that complements the theoretical work of Sulitzeanu-Kenan and Hood (2010). Sulitzeanu-Kenan and Hood present a comprehensive review of the current state of research on political blame and blame avoidance. They propose a general model of blame, followed by a typology of blame avoidance in general, and blame avoidance strategies in particular.

The research question motivating this thesis is: What are the structural links underlying the relationship between social blame-makers and blame-takers? (Hood, 2010) We are also interested in introducing social and political constructs into the study of blame in politics, thus bridging a theoretical gap still present in the current body of research.

To provide a [partial] answer to our guiding question, this work advances a formal model of political blame attribution that describes the relationship between individuals, and agents who are blamed for their actions by these individuals. In doing so, the model attempts to capture how individual blame attribution is transformed into a notion of aggregate blame that is experienced by agents. The model provides an analytical perspective that leans on previous psychological descriptive models of blame formation (Alicke, 2000, Kahneman and Sunstein, 2007, Schlenker, 1994) and integrates their constructs, thus extending them in descriptive resolution and generalizability. Consequentially, it presents a framework within which blame avoidance strategies by these agents can be modeled and analyzed.

The model utilizes and extends a theoretical definition of blame as an interaction between notions of Perceived Loss and Agency Dimension, presented by Sulitzeanu-Kenan and Hood (2005). It elaborates on these concepts and presents formal definitions of several other terms commonly used in the literature on blame and blame avoidance which so far have suffered from inconsistency in use and description. By doing so we are able to draw testable predictions on the outcome of social blame and offer a classification of social blame scenarios.

Following the advancement of the model, we conduct a modest and preliminary exploration of the possible connection between two prominent constructs of political science research – namely, institutional design and social cleavages – and the way social blame is shaped. To accommodate this discussion we extend the theoretical framework of our model by introducing the concepts of coherence and agreement, and provide formal measures for these concepts based on our definitions of social blame. These concepts are then utilized to offer a basic typology of social blame scenarios based on the interaction between coherence and agreement, and their underlying political and social determinants. This typology serves as a potential guideline for future research.

The thesis is arranged as follows: First, the theoretical and methodological background underlying the model is presented. Next, a presentation and discussion of the static version of the social blame model is given. The third section promotes the idea of a causal connection between institutional design, social cleavages and social blame. The fourth part introduces the concepts of agency coherence and agreement, proposes respective measurements for them, and is followed by a typology of social blame situations based on the interaction between these two measures. We conclude with a summary and an outline of possible future developments in the model that incorporate time factors and the implementation of blame avoidance.

2. Theoretical Background

This chapter reviews theoretical and empirical works on blame that constitute the foundations on which our model was developed. The review is largely based on the literature surveyed by Sulitzeanu-Kenan and Hood (2010). We begin with a discussion of the concepts of blame and responsibility. We then review three important theoretical contributions to the understanding of blame formations. This is followed by a focus on affective elements of blame that are constructs in our model.

Blame and Responsibility

The term 'blame' is defined as [to] "consider or say that somebody is responsible for something done (badly or wrongly) or not done"; "be responsible for something bad; deserve to be blamed"; "responsibility for something done badly or wrongly", and "criticism for doing something wrong" (Oxford Dictionary 1989). This set of definitions points out two elements in the notion of blame. First, that it has to do with "something bad" or "wrong". And second, it links this "bad thing" to the responsibility of "somebody". Blame is the act of attributing a 'bad' or 'wrong' thing to a particular person or entity.

It is important at the outset to discern the meaning of 'blame' from that of 'responsibility'. The term 'responsibility' may refer to both positive and negative consequences, and can be both retrospective and prospective[1]. Blame, however, is only associated with adverse events and

[1] The term 'responsibility' holds a wide range of meanings. H.L.A. Hart (1968) distinguished between four senses of 'responsibility': (1) role; (2) causal; (3) liability; and (4) capacity.

consequences. In this respect it is of a narrower scope than responsibility. Yet, for the purpose of discussing the concept of blame, lay responsibility evaluations (as oppose to legal or ethical) for adverse experiences can be used interchangeably with blame evaluations. To the best of our knowledge, all psychological studies of blame and those concerning responsibility as an evaluative judgment of an *adverse* experience, refer to these concepts interchangeably, either explicitly or implicitly. This can be seen by the measurement instruments employed, and their results. Such measurements typically include the term 'responsibility' and 'blameworthiness' in the same measurement scale (Schlenker et al. 1994), and both concepts often employ 'punishment-worthiness' as an item in their measurement scale (Hagiwara 1992, Gibson and Gouws 1999; see also Bartling and Fischbacher 2008 who employ a measure of responsibility that relies on the assignment of actual punishment). These studies suggest that items tapping responsibility, blameworthiness, and punishment-worthiness evaluations, are strongly correlated, suggesting that responsibility for adverse experiences and blameworthiness are strongly associated in people's mind.

Still, legal prescriptions may be the basis for responsibility even in the absence of moral or social blame, while blame may arise in circumstances in which formal responsibility do not. The latter occur when people are evaluated as transgressing moral or social norms, beyond the pale of civil or criminal responsibility (Alicke 2000). For this reason, for the purpose of this model, blame is a more relevant concept to social and political life than responsibility.

A Triangle Model of Responsibility

Schlenker et al. (1994) review of the philosophical and psychological literature identifies two broad facets of responsibility. The first is *imputation*, which features (1) a sense of responsibility as causality; (2) responsibility as a mental state, or more specifically the exercise of free will; and (3) a "mental capacity for acting in a reasoned and deliberate way". The second facet of responsibility is *answerability*. It pertains to (1) the sense of being accountable to others; (2) obligations created by moral or legal codes; and (3) duties arising from social roles (Schlenker et al. 1994: 634).

Relying on these notions of responsibility, suggests that since "responsibility is a necessary component of the process to holding people accountable for their conduct", accountability, as an evaluative reckoning requires information about three elements: *prescriptions*, *event*, and *identity* images (Schlenker et al. 1994: 634). *Prescriptions* are codes or rules for conduct. They furnish criteria for what the actor should be doing in a particular situation; they can serve to guide behavior and to evaluate it. *Events* are the units of action, and their consequences, that actors and observers regard as a unified segment for purposes of some evaluation. *Identity* images refer to the actor's roles, qualities, commitments, and pretensions; they are the components of the actor's identity that are relevant to the situational context (Schlenker et al. 1994: 635).

According to Schlenker et al. (1994) "[a]ll evaluations involve information about all three elements, even though some of this information may be tacitly assumed by the evaluator" (p.

635).[2] The evaluative reckoning involves both the *potency* of the elements, and the combined *strength of the linkages* between them. Prescriptions are more potent when they are regarded as valued principles of conduct; events are important when they pertain to important prescriptions or themselves produce greater consequences"[3]; the identity element is more important when the relevant images are central and valued components of the actor's identity (Schlenker et al. 1994: 636). This distinction between the potency of the elements and the strength of the linkages among them resembles the distinction between perceived loss and agency dimension elaborated later in this paper. However, Schlenker et al. empirical work addressed only the linkages among the elements, and consciously avoids experimenting and theorizing on the effect of outcome on responsibility judgments[4].

---

[2] These implicit assumptions are influenced by affective elements such as spontaneous evaluations (Alicke 2000), and implicit expectations (Kahneman and Sunstein 2007), which are discussed in the next section.

[3] The first part of this definition fails to distinguish between the potency of the event and the combination of the *Event—Prescription* link and the potency of the prescription. However, this flaw does not derogate from Schlenker et al. findings, since their empirical work did not test the consequences of the potency of the elements, only the *strength* of the *linkages* between the elements.

[4] Schlenker et al. (1994) Make a distinction between the notions of responsibility and blame, unlike our view that they are essentially equivalent, in particular when addressed in the social/political context. Their experimental results show that subjects were less inclined to choose information regarding event consequences when asked to form responsibility judgments. Based on these findings, they argue that "prior studies that suggest that consequences information can sometimes affect responsibility may have assessed blameworthiness rather than responsibility per se" (1994: 647-8). We believe that although participants in the described experiment did not try to obtain consequence information for their responsibility judgments, in reality this information is almost always available and is highly pertinent in consolidating responsibility judgments. This is evident in the extensive literature on hindsight bias, as formulated by Fischoff (1975).

Schlenker et al. (1994) main empirical contribution thus relates to the idea of a link between elements. This idea suggests that two concepts are associated, in a way that the evaluation of one generalizes the other. When two elements are linked, they are perceptually associated so that the characteristics of one element come to be associated with the other element. The elements tend to be associated in memory in a manner that when one is recalled, the other is also likely to be recalled. Schlenker et al. thus propose that "the combined perceived strength of the three linkages among the elements determines how responsible the actor is judged to be on the occasion" (p. 635).

Schlenker et al. (1994) experimental findings support their model. The combined strength of the linkages was found to affect responsibility judgments, and information regarding the strength of the linkages was sought more than other types of information for the process of producing responsibility judgments.

Culpable Control Model

Another psychological model of blame is the 'culpable control model' (Alicke 2000). Control in this model refers to the freedom to "effect desired behaviors and outcomes or to avoid undesired ones" (p. 557). Blame evaluations according to this model are constructed by assessing the relationships among three basic elements of behaviors: (1) mental states; (2) behaviors; and (3) consequences. The mind-to-behavior link – *volitional behavior control* – refers to the extent in which a person's actions were freely chosen, rather than compelled; the behavior-to-consequence link – *causal control* – relates to the actor's unique impact the on harmful consequences; and the mind-to-consequence link – *volitional outcome control* – indicates whether the consequences were desired, and whether they occurred as anticipated. "Factors that establish personal control intensify blame attributions whereas constraints on personal control potentially mitigate blame" (p. 557).

Volitional behavior control varies based on the degree to which actors are perceived to have behaved purposely and knowingly. This assessment is based on whether the activity appears to be deliberate rather than accidental, and that the actor seems to be able to understand the meaning of her actions (p. 559). Some information and skills are expected from all adults, while others may be relevant only to people in particular roles (p. 560).[5] Volitional behavior control is diminished by capacity and situational constraints (p. 560).

Causal control reflects the actor's impact on the harmful outcomes (p. 561, see also: Bartling and Fischbacher 2008: 17). The evaluation of causal control depends on the uniqueness of the actor's contribution among other causal factors, but also on whether it was sufficient to produce harm. The second type of considerations includes the temporal or spatial distance between a causal factor and its consequence. A third consideration in assessing causal control draws on counterfactual reasoning, as observers also care about the consequences that would have occurred without the actor's intervention (p. 561).

The third structural element of the culpable control model is 'volitional outcome control', or the extent to which consequences were desired and foreseen. This element is partially dependant on the same situational considerations that affect volitional behavior control and causal control. However, the assessment of foresight may explicitly depart from the actual to the normative,

[5] This variance in expectations is at least partly addressed by the prescription—identity linkage of the 'triangle model', which is absent in this model.

when observers believe that an actor *should have* anticipated a particular consequence (Alicke 2000: 562; This may also be addressed by Schlenker at al. (1994) prescription—identity link.). Observers may also consider process control, or “the process by which constraints evolved”. Put simply, “people can be blamed for relinquishing control” (ibid).

Alicke (2000) volitional behavior control (mind-to-behavior) and causal control (behavior-to-outcome) are covered by the identity—event link in the ‘triangle model’ (since ‘event’ in this model covers both behavior and `outcome). The volitional outcome control is not explicitly part of the Schlenker et al. (1994) model, yet its strength appears to be governed quite substantively by prescriptions (which are absent in Alicke’s model).

These structural models of blame thus suggest that the strength of the mental linkages between elements is determined by assessing the control – whether actual or normative (prescribed control), and the application of information regarding the relevance of prescriptions to identity and events.

Aggregate Responsibility Formation

Bartling and Fischbacher (2008) explore the relationship between responsibility attributions, blame expressed by punishment and the delegation of responsibility. They attempt to explain who is held morally responsible for the outcome of a delegated decision. Following this motivation, they present a measure of responsibility perceptions that treats player responsibility as the aggregation of individual beliefs regarding a player. In the model underlying the measure, individuals carry beliefs on players’ strategies by which they infer the conditional probability of an outcome. To compute the level of responsibility attributed to a specific player by the participating population, the different inferred responsibilities for outcomes associated with the player are integrated over the distribution of beliefs in the population. The resulting formula therefore amalgamates personal attitudes into a single notion of social responsibility.

In a different part of this work, Bartling and Fischbacher (2008) study the effectiveness of delegation as a blame avoidance strategy. They conduct an experiment based of the dictator game, where some subjects are required to allocate a certain amount of points in a 'fair' (equal funds for all) or 'unfair' (giving more points to the allocating players over receiving players) manner, and under some conditions have a choice of delegating the decision right to a second allocator, who in turn chooses between the fair or unfair distribution. The subjects on the receiving end of the dictator game can then decide whether or not to punish the delegator, the delegee, or both by depriving them of a part of their allocated funds. The punishment patterns are used as a measure of the receiving subjects' attitude towards the allocators. This design measures blame attributions using the strength of punishment inflicted upon players and therefore identifies blame with the expression of discontent.

Bartling and Fischbacher then demonstrate that their measure of responsibility better explains the variation in punishment (i.e. blame) patterns present in the experiment when compared to an outcome-based measure. By linking aggregate responsibility formations with the prediction of blame patterns, they not only explain in detail where moral responsibility lies when a decision right is delegated, but they also offer a theoretical framework that allows for a definition of blame based on public blame attributions.

Some other important design considerations taken by Bartling and Fischbacher are that no single individual is more 'powerful' (i.e. has a stronger social effect) than the others in forming responsibility judgments, and that the overall responsibility attribution is bounded and thus constitutes a zero-sum game by the involved blame-takers. They assert, regarding the responsibility for the probability of an outcome, that "*if more than one player increases this probability, then each player's share in the overall probability increase is calculated*" (2008: 18).

Affective Elements of Blame

Blame is also shaped by other, less rational, normative, and calculated considerations. Schlenker et al. concede that "Different audiences may, of course, differ in their assessments of the relevant prescriptions, events, and actor's identity (p. 635). This variation, left unaccounted for in the structural models is more directly addressed by the literature on spontaneous and affective reactions. These elements of blame are discussed in the following section.

Early psychology writing assumed that unconscious motives and feelings drove people's judgments, while reasoning played an ex post role of communicating the judgment in a socially acceptable way (Freud 1900/1976, as quoted by Haidt 2001: 816). Yet for a long period a rationalist conception of morality as a system of abstract rules that can be understood and internalized (Kohlberg, 1969) has dominated moral psychology. More recent studies of moral judgment have redirected attention to the role of spontaneous emotional reactions that are not anchored in reason.

Such an approach relates to the wider notion of dual-process theories, which "distinguish cognitive operations that are quick and associative from others that are slower, more reflective, and frequently more calculative" (Kahneman and Sunstein 2007: 4). It is suggested that moral judgment is initially shaped by the first process.[6] Such spontaneous evaluations may also be (re)considered in a reflective and reasoned way, but this latter process is not independent from the outcome of the preceding one. The existence of both processes may lead to 'moral dumbfounding' (Haidt et al. 2000) – an experience of strong moral reactions for which no adequate reason comes to mind, and 'moral numbness' – "in which people are not indignant even though they have reason to be, and know they do" (Kahneman and Sunstein 2007: 2).

Spontaneous evaluations are particularly relevant to blame in the political context. In contrast to criminal and civil offenses, which are governed by legal codes and case law, social and political wrongdoings are typically contentious and plagued with widespread disagreement about their nature and severity. "Observers' personal attitudes and values, therefore, have considerable latitude to influence their interpretations" (Alicke 2000: 567). The subjectivity inherent in social offenses facilitates the influences of spontaneous and emotional evaluation on judgment.

Perceived Loss and Agency Dimension

Perceived loss and *agency dimension* are suggested as two organizing concepts of blame formation that draw on the preceding sources. By introducing these two concepts, we draw on Felstiner et al. (1980). In their work, blame is preceded by the notion of "injurious experience", defined as "any experience that is disvalued by the person to whom it occurs" (p. 634).[7]

[6] This relies on evidence, both behavioral and neurophysiological, that the assessment of whether objects are good or bad is carried out quickly and efficiently by specialized neural circuitry (Kahneman and Sunstein 2007: 6).

[7] This starting point avoids more morally loaded terms, such as 'bad' and 'wrong', and establishes the conception of blame on subjective judgment, allowing a descriptive rather than normative analysis of blame.

However, a mere conscious injurious experience is not enough. “(I)n order for disputes to emerge and remedial action to be taken, an unperceived injurious experience must be transformed into a perceived injurious experience” (p. 633). Felstiner et al. refer to this first transformation – “saying to oneself that a particular experience has been injurious” – as *naming* (p. 635). The next step – *blaming* – occurs when a person attributes an injury to the fault of another individual or social entity (ibid). This notion of blame suggests that blame is the result of two elements – a perceived negative experience, termed here *perceived loss*, and an *agency dimension* of blame, that results from attributing and relating the experience to a particular agent (see: Alicke 2000: 558;Brändström and Kuipers 2003; Bovens et al. 1999: 126; Schlenker et al. 1994 p. 633; see also 'traceability' in Arnold 1990).

How individuals judge both perceived loss and agency dimension in a given situation involves spontaneous evaluations that balance between prescriptions – the assignments of a unique status to a particular reference state – and a perceived outcome, which provides evidence of deviations from the prescriptions (Kahneman and Sunstein, 2007: 7-10; also Schlenker et al. 1994: 634). This applies to both the negative experience – i.e. the perceived loss – and the different agency dimension attributions associated with that experience.

Agency Dimension and Blame Validation

Emotional reactions take various forms. One notable process is a variant of ‘attribute substitution’ (Kahneman and Frederick 2002), a process by which people faced with a difficult evaluative task (e.g. determining the appropriate punishment) often conduct an easier one instead without being aware of the substitution. In the case of moral evaluation they may resort to consulting the intensity of their outrage - a process termed ‘outrage substitution’ (Kahneman and Sunstein 2007: 7).[8] Other spontaneous reactions occur in response to both evidential (structural linkage elements) and extra-evidential information – factors such as class, race, gender, or social attractiveness. These reactions activate the desire to blame the person or persons who evoke the most negative affect (Alicke 2000: 564). This is attained by altering evidential standards, influencing control perceptions, or leading observers to search selectively for information that supports a desired blame attribution. Together, these amount to a ‘blame-validation processing’ – “the proclivity to favor blame versus nonblame explanations for harmful events and de-emphasize mitigating circumstances” (p. 568).

Blame-validation processing is enhanced by the tendency to view people rather than the environment as the controlling forces behind harmful events. Correspondence bias literature suggests that human agency attributions are favored over explanations that involve mitigating circumstances, are less modifiable once they are formulated, and readily supersede environmental hypotheses that mitigate blame (Alicke 2000: 568; a number of psychological reasons for this preference are given on pp. 568-9).

To conclude, agency dimension – the association between a perceived loss and an identified agent – is shaped by assessments regarding the actual and/or prescribed control that the agent has over the perceived loss. Such assessments are influenced by spontaneous and affective reactions, which tend to bias assessments of control and motivations, and by a proclivity to seek the human factor and favor blame attributions when experiencing loss.

[8] William Ewart Gladstone (1809-1898) is credited for similarly arguing that “Men are apt to mistake the strength of their feeling for the strength of their argument”.

Perceived Loss and Prescriptions

Perceived loss is a negative experience that is recorded consciously as such. Greater and more potent losses evoke stronger blame attitudes when coupled with agency dimension (Schlenker et al. 1994, Kahneman and Frederick 2002: 64). Although losses are typically grounded in factual outcomes, their identification and perception as such are also the product of mental processes. These processes, like other evaluative tasks, are susceptible to biases that stem from emotional reactions and cognitive limitations. Understanding some of these processes provides a clearer basis for analyzing and evaluating blame.

The weight given to perceived loss in shaping blame evaluations is such that it tends to bias such evaluations, producing moral dumbfounding. For example, the moral difference between an intentional reprehensible act that accidentally failed to cause harm (e.g. attempted murder), and one such act that produced harm is questionable. The emotional response that is elicited by the gravity of the loss acts to shape the intuitive evaluative process, while the evaluator may not be able to reason her judgment when confronted with the two situations (Kahneman and Sunstein 2007: 9).

Some losses may be unequivocal, as Hobbes' *summum malum* – a premature and violent death – the fear of which Hobbes identifies as the fundamental social motivation (Flatham 2002: 76). Yet, the attribution of blame is not a reflexive response to objective events. Rather, blame is assigned by filtering events through a perceptual screen (for references, see Ellis, 1994: 5, f.n. 21). Quite often factual outcomes become perceived as harms or losses only against the backdrop of certain prescriptions or entitlements. Such a relationship is captured by Schlenker et al. (1994)

prescription—event link, a norm that is applicable for evaluating conduct and outcomes. Such prescriptions are available in legal and ethical domain, and by deliberate and calculative effort can be applied to produce normative evaluations of behavior and outcomes.

However, this leaves unanswered the question of whether prescriptions play a role in intuitive and spontaneous evaluations, and if so, how. Kahneman and Sunstein suggest that "an entitlement is a socially endorsed *normal state*, also called a *reference state*, relative to which losses are defined". Such a notion of prescription – the assignment of a unique status to a particular reference state (typically, but not necessarily the status quo), and detecting deviations from it, is attainable in spontaneous and brief processing (2007: 10). For example, Kahneman and Miller (1986) have found that fairness judgments regarding economic transactions are guided by a 'reference transaction' that specifies the relationship between the reference profit for the firm and reference terms for the transactors. A change to the detriment of the consumer was not judged as unfair if it was imposed in order to minimize a *reduction* in the profit of the producer. This example suggests that intuitive evaluations (and their accompanying biases) may influence the prescription—outcome linkage, and hence the perception of a loss.

Other biases may also directly affect the judgment of outcomes. Some harms may seem more poignant when their counterfactual 'undoing' comes more easily to mind. For example, an outcome that came about in the course of an unusual turn of events would be judged as more poignant than an equivalent outcome of a common occurrence (Kahneman and Sunstein 2007: 7).

It should be emphasized that although prescriptions readily lend themselves to the theoretical description of perceived loss via the prescription-outcome link, this relationship also serves to describe the underlying mechanism of agency evaluations. Individuals maintain prescriptions not only relative to an event but also relative to the behavior of agents, thus allowing for a perception of an outcome to be judged relative to both event prescriptions and agency prescriptions, affecting both the formation of perceived loss and agency blame attributions.

Our conceptualization of outcomes (as presented in the next chapter) describes them as having values that are equal or lower than the value given to the prescriptions governing the situation. This reflects the idea that the blaming individual is incapable of experiencing 'positive surprise' by an outcome. This design therefore does not accommodate the formation of credit attributions (i.e. negative blame values) by individuals. We make this choice because our model aims to describe the realm of blame attributions rather than attitudes in general. Therefore, any observation of behavior that is in accordance, or exceeds, the relevant prescriptions is only taken to indicate a lack of negative attribution (although it can certainly have other repercussions that are not within the scope of this model). This modeling decision draws on the observation that blame and credit attributions are not mirror reflections around an axis of symmetry, which would have allowed for a measure of outcome that exceeds the prescription, and in turn, creates a negative perceived loss value. Instead, they are governed by biases similar to the ones that characterize utility perceptions in Prospect Theory (Kahneman and Tversky, 1979) in the sense that while negative outcome judgments tend to be exacerbated relative to the 'true' damage caused; positive outcomes produce less credit attributions (in similar vain to 'blame validation', see: Alicke, 2000: 558). This design choice is also in line with empirical results presented by McGraw (1991: 1143) where credit and blame attributions for the same agents were shown to be independent.

Coherence and Agreement

This part introduces the two theoretical concepts of coherence and agreement, which are formally developed and defined in chapter 5.

Blame can vary in its degree of abstraction, in a spectrum that, on one dimension, runs from blaming 'capitalism' for 'the class system' to blaming a particular person for a specific injurious act. On another dimension, it runs from blaming a multitude of actors to blaming a specific agent as the sole person responsible for a perceived loss. Furthermore, the level of consistency within a specific population on who is to blame can vary from full agreement between every two individuals on who is to blame to many contrasting views as to the actors responsible for an injurious event.

It is suggested here to introduce two aspects that are pertinent to the description of blame in social (rather than interpersonal) settings, namely the *coherence* of the agency dimension – how concentrated is the attribution of blame to particular agents out of a range of possible agents – and the *agreement* within the population – whether the blame distribution attained in a situation is a result of a shared common opinion or an amalgamation of different blame attributions that together form the aggregate social blame.

These features of agency dimension extend the psychological elements discussed so far, and justify the conceptual distinction between the terms 'attribution' and 'agency dimension'. Agency dimension, as a concept that denotes the strength of the association between a perceived loss and an agent, encompasses attribution, but also the variability that results from different *distributions of attributions* among varying sets of agents.

When a given perceived loss is equally associated with a number of agents, agency dimension can be said to be incoherent, as no clear net vector of blame can be derived. Agency dimension thus extends the possibilities to weaken the prescription-identity link (Schlenker et al. 1994) by introducing situations of multiple identities – thereby reducing the *individual* link of each agent.

The consequences of agency coherence on the emergence of blame had not been extensively studied in political science. Yet some researchers provide support to the hypothesis that agency dimension coherence is positively related to the emergence of blame. Wilson (1961) had proposed that the passive response of black Americans to widespread problems in housing is in part a result of the large and confusing number of culprits and potential remedies for the problems. Javeline (2003) has found in a nationwide survey of the Russian population that "the greater the specificity of blame attribution, the greater the probability of protest", i.e. that specificity of attribution is positively linked to the expression of blame. Javeline has suggested that specificity may reduce the cost of information, organization and opportunity, associated with effective public protest (Javeline 2003: 109).

Another strand of the literature lends, less direct, yet reasonable support for the importance of agency dimension coherence. Comparative analyses of economic voting in Europe and in the American states suggest that the economy—vote/approval relationship is attenuated in countries (or states) with divided government or with high levels of coalition complexity (Anderson 1995, 2000; Lewis-Beck 1988; Leyden and Borrelli 1995; Powell and Whitten 1993; Rudolph 2003a). Assuming that institutional structures affect agency dimension coherence, these findings support the relationship described above between agency dimension coherence and blame.

Further research is needed regarding the effects of agency dimension coherence on blame, particularly as some blame avoidance strategies are explained on the basis of reducing agency dimension coherence (e.g. ‘circle the wagons’ (Weaver 1986), and ‘protocolisation’ (Hood 2004)).

While the concept of coherence did receive some (indirect) attention, the notion of agreement and its effects on social blame is left virtually unexplored. Javeline (2003) has found that agency dimension coherence at the individual level is linked to the propensity to protest, yet data on the effects of aggregate level agency dimension agreement on blame expression is lacking.

3. A Formal Model of Social Blame

This chapter presents a formal model of social blame formation in a multiple-agent, multiple-individual environment that is based on the constructs of blame surveyed in Chatper 2 and on past works that tried to integrate them. We first elaborate on some methodological considerations and then follow with a serial introduction of the definitions that constitute the model. We conclude the chapter with a discussion of several issues arising from our design and its interaction with standing work and data.

Methodological Considerations

What structural links constitute the relationship between social blame-makers and blame-takers? In Alicke (2000: 556) review of blame research he notes that most work in the area tended to focus on specific facets of blaming[9]. However, he also concludes that few clues have been provided on the interrelationship among the constructing components of blame or on the process by which blame attributions are made.

This observation becomes more pronounced when we consider blame attributions in a multi-agent, multi-individual environment – namely, cases of social or political blame attribution. Formally describing the different actors and the constructs of their blame attributions, and further

[9] According to Alicke, these include the motivational assumptions of blame attributions; psychological issues (such as eyewitness identification and jury selections); effects of outcome information on blame ascriptions; how blame or responsibility attributions are influenced by different characteristics of the observer, perpetrator, or victim; and how specific criteria (causal impact, foresight of the consequences, intervening causation, intention and motive, mitigating circumstances) are used in blame ascriptions

elaborating the interaction between these actors and these factors, allows for a better theoretical understanding of how aggregate blame is formed and directed toward specific agents. It also produces testable predictions on the behavior of blame-taking agents and blame-making individuals, and further allows for a theoretically grounded incorporation of blame avoidance strategies and outcomes within it.

The formal model of blame advanced in this thesis suggests that individual blame attribution is the product of the cognitive linkage between a perceived loss and an (which consequently becomes "its") agency dimension. In line with Felstiner et al. (1980), and based on a number of more recent studies in social and political psychology, it is maintained that since individual blame attribution is a *directional* attitude, it is not independently linked to the *size* of the perceived loss, but rather the result of an interaction with the *direction* of agency dimension (for a related example of how such interaction affects the domain of voting behavior, see Arcenaux, 2003[10]). Formally speaking, the blame an individual i attributes a particular agent a regarding perceived loss on an issue s is a function of the potency of this individual's perceived loss multiplied by the strength of the association between agent a and the perceived loss on issue s (see: Kahneman and Frederick 2002: 64).

It is further suggested that the notion of individual blame is aggregated into a concept of collective blame, which represents how agents perceive the amalgamation of individual blame attributions as their conception of blame over the entire population they consider relevant

[10] Arcenaux demonstrates how economic adversity – akin to perceived loss – and government blame attribution with regards to the economy – the agency dimension – have a compound effect on voter turnout; The presence of both an adverse effect of the economic situation and a tendency to see the government as baring responsibility for the economic hardships is a necessary for increase in voter turnout.

Design

The model assumes that the population is composed of two types of people: *individuals* and *agents*. Agents are, generally speaking, actors who are involved (or perceived to be involved) in a situation either by acting in a manner that affects it, or by being morally, normatively, institutionally or lawfully responsible for certain aspects of that situation and its outcome. Individuals, on the other hand, have no direct influence on a situation and its outcome, but instead perceive themselves to be (negatively) affected by it. Individuals act by attributing blame to agents after they have experienced some degree of damage (actual or symbolic) in a specific situation.

Agents in the model are assumed to be rational in the sense that they act in a manner consistent with their goal of minimizing the amount of blame attribution they perceive, and that they are completely aware of public blame attributes directed at them and at the other agents (in line with the theoretical framework laid out by Stimson, MacKuen and Erikson, 1995: 544-545, and consistent with data in Burstein, 2003, Page and Shapiro, 1983, and Weaver, 1986). In contrast, we do not assume that individuals carry any rational attributes of utility maximization or strictly conscious choice among their alternative courses of actions. Instead, individuals are assumed to attribute blame based on a mixture of spontaneous evaluations and elaborate reasoning (demonstrated by Kahneman and Sunstein 2007, and discussed earlier).

Definition 1: Let I denote a set of *individuals,* marked $i \in I$.

Definition 2: Let A denote the set of agents. An *agent* $a \in A$ is responsive to the weighted aggregate blame directed towards her by the individuals (as defined hereinafter).

Let us further assume that I and A are mutually exclusive ($I \cap A = \emptyset$), such that no individual is also an agent and vice versa. This simplifying assumption eliminates from the model's descriptive scope situations in which the actors are both individual blame attributors and blame receivers. These situations dominate the psychological literature on interpersonal blame (Alicke 2000: 556) but are less relevant for describing the formation of aggregate blame. This is justified on the grounds that in most political situations, $I \gg A$. Therefore, even if $A \subset I$, an agent who is also a blame attributor has only a marginal influence on the overall blame outcome[11]. Furthermore, in political blame formation, the agents' option of "blaming back" the blaming individual (rather than agent) is practically nonexistent.

Definition 3: An *issue s* is any event that involves a perception of loss by at least one individual who is not personally involved in its outcome as an active agent (e.g. a citizen that is affected by a decision to impose a tax but has no direct power to change the decision), and at least one agent, who is perceived by at least one loss-baring individual as having to do with the outcome of that event. An agent can also be involved in an issue owing solely to institutional or normative expectations that cause her to be perceived as present in or associated with the situation, even without having any individual blame attribution directed at her.

---

[11] Agents also have the option to direct blame at other agents, but this concept of blame attribution stems from different constructs. In particular, agents do not experience a perceived loss (the term is elaborated hereinafter) in a manner similar to individuals, nor are they taken into consideration in the same way as individuals when another agent assesses his public blame level. Instead, blaming by agents serves a 'higher' level of impact, affecting overall blame indirectly, thus representing a blame avoidance strategy meant to direct blame away from them and unto other agents. Blame avoidance, although a fundamental construct of social blame and an implicit design consideration in our model, is not covered in this work.

Let S be an arbitrary set of issues and $s \in S$ a specific issue. s therefore identifies a pair (I^s, A^s) consisting of a subset $I^s \subseteq I$ of all individuals i in the population who experience loss by the issue, and a subset $A^s \subseteq A$ of all agents a for whom exist at least one individual who attributed them some involvement in the outcome of the issue, or agents that carry an institutional or normative relevance to the event (e.g. on an issue pertaining to a country's budget, the country's secretary of treasury would still be counted as an agent relevant to the issue regardless of their perceived effect on the issue).

Definition 4: An individual $i \in I^s$ holds a set of *prescriptions* regarding the issue, which is expressed by a value ${P^s}_i$ that encompasses her norms and expectations[12]. These prescriptions are the background against which an individual experiences a specific observed behavior – the perceived *outcome* of the issue, denoted ${O^s}_i$ (Kahneman and Sunstein, 2007: 7-10). We assume further that ${O^s}_i \in [0, {P^s}_i]$ and that if ${O^s}_i = {P^s}_i$, the outcome meets the individual's set of prescriptions regarding the issue, and therefore no blame is generated.

Following this, individual i has *agent-specific prescriptions* ${P^s}_{i_a} \in {P^s}_{i_A}$ that denote his or her norms, expectations and affective predispositions towards an agent's behavior on an issue. They are respectively matched by the individual's observation of each agent's behavior on the issue, denoted ${O^s}_{i_a} \in {O^s}_{i_A}$. Similarly, ${O^s}_{i_a} \in \left[0, {P^s}_{i_a}\right]$. This design is similar to the notion of strategy choice belief presented by Bartling and Fischbacher (2008: 18) and also embodies the causal control criteria formulated in Alicke (2000: 561), and the distinction between prescriptions and consequences suggested by Schlenker et al. (1994: 647-8).

[12] A further discussion of this design choice is provided at the end of this chapter.

Both the prescriptions regarding the perceived outcome of an issue and the prescriptions regarding the perceived agents' behavior on an issue are interrelated and affect each other. Jointly, they constitute an individual's complete set of issue prescriptions.

Definition 5: Drawing on definitions 1-4, the *perceived loss* of an individual *i* is defined as:

$$PL_i = \begin{cases} 1 - \frac{O^s{}_i}{P^s{}_i}, & P^s{}_i > 0 \\ 0 & P^s{}_i = 0 \end{cases}$$

Therefore, $PL_i \in [0,1]$, and the larger the difference between an individual's prescription of an event and the outcome, the closer to 1 will be the value of the individual's perceived loss. In the private case where $P_i = 0$ (i.e. no expectations or norms involved on an issue) we set PL_i to zero, signifying no perceived loss.

In line with the theoretical discussion, in this definition of the outcome $O^s{}_i$, its value cannot exceed the prescription $P^s{}_i$. We stress that, meeting the considerations noted in the previous chapter, an outcome value of zero does not entail a behavior that is the complete inverse of the prescription. Instead, it signifies any behavior that completely fails to meet the normative framework or expectations of an individual regarding an event. In this respect, an outcome which is orthogonal to the individual's prescription set may also carry a value of zero by merit of not meeting the prescriptions, despite not being directly opposite to them. The assignment of a zero value, therefore, is practical in nature and is not suggestive of a bounded spectrum of behaviors relative to specific prescriptions.

Definition 6: The *agency dimension* of an individual *i* relative to an agent *a* is defined as:

$$AD_{i_a} = \begin{cases} 1 - \dfrac{O^s{}_{i_a}}{P^s{}_{i_a}}, P^s{}_{i_a} > 0 \\ \quad\quad 0 \\ 0 \quad\;, P^s{}_{i_a} = 0 \end{cases}$$

and thus carrying the same properties of perceived loss only with regards to the outcome of an agent's perceived behavior within an event, rather than the general outcome(s).

Definition 7: The *individual blame attribution* of an individual *i* towards an agent *a* on issue *s* is defined as:

$$b_{i_a} = PL_i \times AD_{i_a}$$

Figure 1 graphically displays this interaction by showing that the agency dimension is the vector that directs the individual's perceived loss towards the agent:

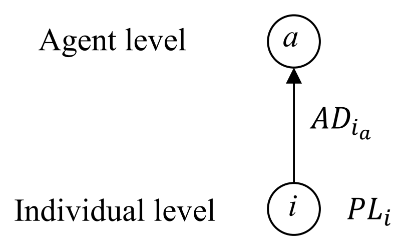

Figure 1 - individual blame attribution

Definition 8: Assume there is a single agent *a* involved in an issue *s* (i.e. $|A^s| = 1$). We define the agent's *aggregate blame* to be:

$$B_a = \sum_i b_{i_a} = \sum_i PL_i \times AD_{i_a}$$

Where the summation is on all $i \in I^s$. This notion is depicted in Figure 2, where $I^s = \{i_1, i_2, i_3, i_4\}$:

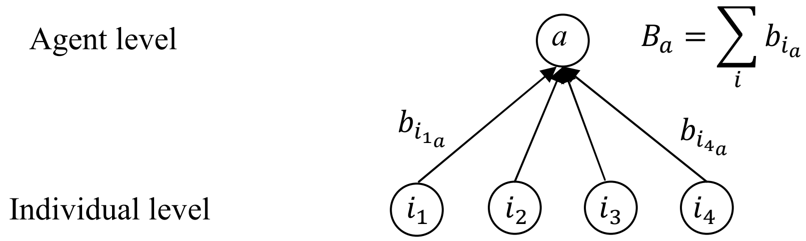

Figure 2 – aggregate blame

This definition of aggregate blame represents how an agent perceives the accumulative blame attribution by individuals on an issue. As such, an agent's perception is susceptible to alterations due to innate biases and, perhaps more importantly, due to reliance on indirect information channels (media coverage, policy advisors, public opinion polls, anecdotal evidence). Despite that, we make the assumption that even in the presence of these biases, agents are able to draw an accurate picture of the aggregate blame attributions towards them, and, more importantly, that they act upon this information. This assumption is supported by extensive empirical findings on politicians' response to public opinion, which demonstrate that in most cases, and especially as issue salience increases, public figures tend to adjust their actions and policy choices according to public opinion (for an overview see: Burstein, 2003, Page and Shapiro, 1983).

By making this assumption we do not claim that the aforementioned biases are negligible, but rather that taking them into consideration does not carry explanatory benefits in our model[13].

---

[13] Further development of the model and inclusion of more diverse patterns of behavior (especially blame avoidance strategies that affect how agents construct their self-image) may require an explicit inclusion of a masking function that expresses more subtle effects of blame impression formation by agents.

The definition of aggregate blame as a measure of agents' perception of their social blame attribution, and the further derivations that follow from this definition, are also in line with Bartling and Fishbacher (2008: 18-20) aggregate definition of responsibility.

We now relax the assumption that there is only one blame-taking agent involved in an issue, allowing each individual to attribute blame to more than one agent.

Definition 9: Assume $|A^s| > 1$. Therefore an individual *i* has a vector of individual blame attributions $\begin{bmatrix} b_{i_{a_1}} \\ \vdots \\ \vdots\ b_{i_{a_n}} \end{bmatrix}$ that relates to all or part of the agents relevant to the issue *s*. Let $\sum b_{i_a}$ denote the sum of individual *i*'s blame attributions over all agents $(b_{i_{a_1}} + b_{i_{a_2}} + \cdots + b_{i_{a_n}})$. The share of blame attributed by individual *i* to agent *a* from *i*'s overall vector of blame is the *weighted individual blame attribution* on individual *i* towards agent *a*. It is defined as:

$$wb_{i_a} = \frac{b_{i_a}}{\sum b_{i_a}}$$

This definition scales the individual blame attributions values to the segment $[0,1]$ but retains their relative proportion. In a private case where there is only one individual, we can use the following for representation:

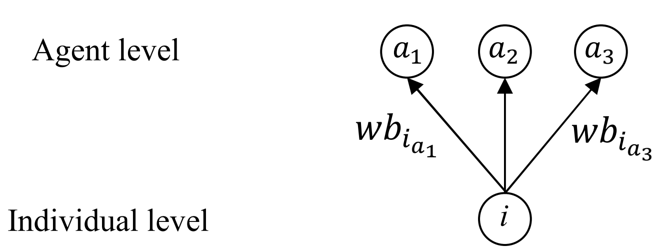

Figure 3 – weighted individual blame attribution

Note that since this scaling process is carried out on all participating individuals, it results in every individual having the same effect size on aggregate blame[14]. This expresses an underlying assumption in the model that no single individual is more 'powerful' (i.e. has a stronger social effect) than the others. This design implicitly governs many accounts of political blame and blame avoidance (e.g. Bartling and Fischbacher 2008, Hood 2002, McGraw 1991, Weaver 1986), and it clearly becomes more plausible as $|I^s| \to \infty$ .Yet, we posit that this assumption also holds for smaller groups up to a certain size, as long as they are large enough so that agents do not have a personal acquaintance with most of the individuals (which would make them inclined to consider the familiar individuals' blame attribution more strongly), and so are motivated to take into consideration a broader, more balanced picture of blame attribution.

Having introduced multiple agents, we now expand on definition 8 and define their notion of social blame in relation to the weighted individual blame attributed to them by the individuals on an issue.

Definition 10: Assume that $|A^s| > 1$. Then an agent a's *aggregate blame* on issue s is defined as[15]:

[14] Different individuals can of course have different blame attributions towards different agents, but the sum of their weighted individual blame is scaled down to one, which means that every individuals has the same 'strength' in affecting the aggregate blame formation.

[15] This definition of aggregate blame holds for the common case where there is more than one agent taking blame on an issue. In the private case where there is only one agent in play, we revert to definition 8 and sum over the non-weighted individual blame attributions. This definition is chosen because where there is only a single agent, her weighted blame attribution will always amount to one, losing any descriptive significance. By summing over non-

$$B_a = \sum_i wb_{i_a}$$

where the summation is on all $i \in I^s$.

In a multiple agent environment the size of the aggregate blame value has no meaning on its own, but rather attains significance when compared to the other agents' blame attributions. (see Bartling and Fischbacher (2008), who suggest a similar, zero-sum, distributive notion of responsibility for the probability of an outcome). To achieve a notion compatible with this purpose, we bound and scale down each agent's aggregate blame in proportion to its relative part of the total blame, as follows:

Defintion 11: Assume $|A^s| > 1$. The *weighted aggregate blame* of agent *a* on issue *s* is defined as:

$$wB_a = \frac{B_a}{\sum_a B_a}$$

where the summation is on all $a \in A^s$.

This weighted design of aggregate blame emphasizes the relativity of blame when multiple agents are involved. This concept and the distinction between aggregate blame and weighted blame also allows us to further develop the concepts of coherence and agreement in the next chapter.

---

weighted attributions in the single-agent case, we allow for variation in aggregate blame strength. This enables a comparison of blame levels attained by the agent when conditions change (e.g. the strength of blame attributions by some individuals is increased / reduced).

A graphical representation of this final definition appears in Figure 4:

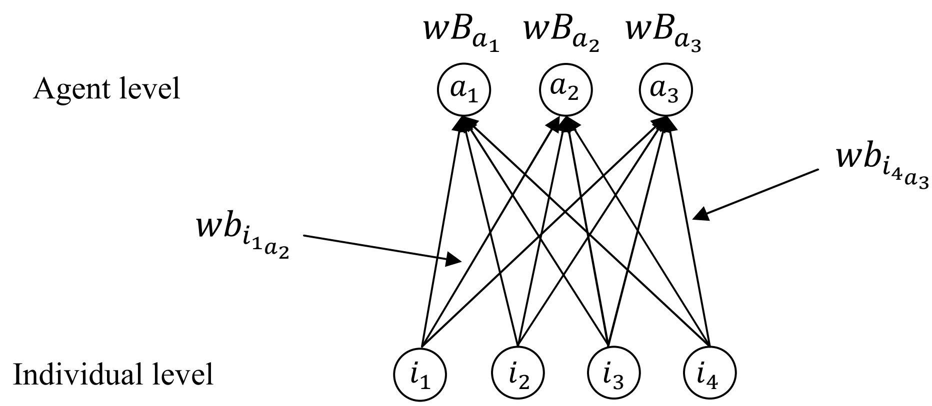

Figure 4 – weighted aggregate blame

An important design choice in this model is that every agent considers blame attributions made by all of the participating individuals. This entails the view that when judging for blame, all politicians give equal consideration to all groups and sectors of the public, regardless of their different effect on politicians` interests, such as reelection chances (in contradiction with Weaver 1986). Indeed, in reality it is easy to conceive that an agent may only be sensitive to blame attributions made by specific subsets of the population (e.g. a politician's constituency, party members, different pressure groups, etc.) and intentionally ignore the rest of the public's sentiments. This choice therefore blurs a (possibly major) behavioral distinction that may be a strong factor in how agents respond to social blame – especially when crowds are heterogenic with respect to an issue or an agent. In the scope of this work, however, the presented model does not account for this characteristic. Future work would benefit from making this differentiation and from comparing agents' blame and resulting behavior with and without this subtlety.

Discussion and Implementation on Existing Data

This part addresses some issues that arise from linking our model to other contributions to the study of social blame. We first use experimental data obtained by Bartling and Fischbacher (2008) and compare them with the model's predictions in the same scenario. We then discuss some findings by Schlenker et al. (1994) that may be indicative of the constructs of our models. Lastly, we demonstrate the importance of using an interval scale for individual blame attributions for our model's validity.

As described in the theoretical review, Bartling and Fischbacher (2008) study the effectiveness of delegation as a blame avoidance strategy. They conduct an experiment based on the dictator game, where some subjects ("Player A") are required to allocate a certain amount of points in a 'fair' (equal funds for all) or 'unfair' (giving more points to the allocating players over receiving players) manner, and under some conditions have a choice of delegating the decision right to a second allocator ("Player B") who in turn chooses between the fair or unfair distribution. The subjects on the receiving end of the dictator game ("Player C") can then decide whether or not to punish the delegator, the delegee, or both by depriving them of a part of their allocated funds. The punishment patterns are used as a measure of the receiving subjects' attitude towards the allocators.

By simulating the experiment with our model we are able to obtain results similar to those recorded by the researchers[16].

[16] A similar procedure is performed by the researchers themselves, who offer different measures of punishment motivation, separately based on agent responsibility, agent intention and allocation outcome (pp. 21-23), and compare them with the empirical findings.

There are six conditions on which punishment patterns were measured by Bartling and Fischbacher. The first two consist of a no delegation scenario where the allocation is solely made by player A. In the first condition, A chooses an unfair allocation, dividing 20 points by giving herself and player B 9 points each, and leaving the two C players with 1 point each. In the second condition, A chooses a fair allocation, equally allocating the 20 points to herself, player B and two player C's (5,5,5,5).. The remaining four scenarios involve the possibility of delegation. A either chooses not to delegate, therefore making a fair allocation (condition 6 in the paper) or an unfair allocation (condition 3), or she delegates the decision to player B, who in turns allocates fairly (condition 5) or unfairly (condition 4). The conditions are summarized in table 1:

	No delegation	**Delegation possible**	
		A allocates	**A delegates, B allocates**
Unfair allocation	Condition 1	Condition 3	Condition 4
Fair allocation	Condition 2	Condition 6	Condition 5

Table 1 - summary of experimental conditions in Bartling and Fischbacher (2008)

These behaviors are realized using manipulations of a one-individual, two-agents version of our model, as illustrated in Figure 5[17]:

[17] Note that we include only one C player in our simulation. If we had included two player C's, the assumptions regarding their behavior would be identical and not affect the result, thus rendering the second player redundant. This design therefore captures the original setting in the experiment. Bartling and Fischbacher (2008) included two player C's to allow for an equal allocation choice between agents A, B and the receiving end (player C's).

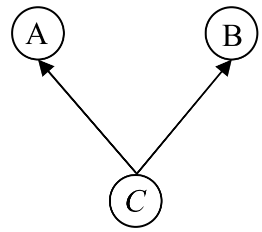

Figure 5 – simulation structure of Bartling and Fischbacher (2008) experiment

We make the reductive assumptions that the two player C's prescriptions only capture their interests with regards to the allocation and the related agent behavior. Perceived loss value is therefore determined by the two player C's expectations from the allocation. We assume all individuals prescribe to an equal allocation, and therefore $P_C = 5$. The perceived outcome is A or B's allocation with respect to player C, which is either $O_C = 5$ (fair allocation) or $O_C = 1$ (unfair allocation). This results in perceived loss values of 0 for a fair allocation ($PL_C = 1 - \frac{O_C}{P_C} = 1 - \frac{5}{5} = 0$), rendering our prediction of blame attributions in conditions 2, 5 and 6 to be zero for both player A and player B.

We now assign the agency prescription and outcome values for conditions 1, 3 and 4, involving unfair allocation. When the allocation is unfair, $PL_C = 0.8$, allowing for blame attributions with positive values.

We make the simplifying assumption that two player C's agency prescriptions are only concerned with the allocation's result and the decision to delegate. When player A can delegate her decision, player C expects her both to make the decisions on her own without delegating, and not to allocate unfairly. Since player B can, at most, decide upon an allocation, player C's prescriptions of Player B concern only the expectation that she will not choose the unfair allocation.

In condition 1, delegation is not possible and therefore player C has an agency prescription value of zero for player B, resulting in player B having an agency dimension of zero ($AD_{C_B} = 0$). Player A is expected not to choose the unfair allocation ($P_{C_A} = 1$)[18] but nevertheless does so, resulting in $O_{C_A} = 0$, and therefore $AD_{C_A} = 1 - \frac{0}{1} = 1$. This translates to predicted individual weighted blame values of 1 to player A and 0 to player B[19] ($wb_{C_A} = 1, wb_{C_B} = 0$).

In condition 3, A has the option of delegating, chooses not to, but makes an unfair allocation, which results in the same blame pattern: B maintains zero prescription value and so $AD_{C_B} = 0$ and her individual weighted blame value is expected to be zero. A has a prescription of 1, consisting of both the expectation not to delegate and to not allocate unfairly. This is only partially fulfilled, which means $O_{C_A} = 0.5$, $AD_{C_A} = 1 - \frac{0.5}{1} = 0.5$, and therefore a predicted individual weighted blame value of 1.

In condition 4, A delegates to B, who in turn chooses the unfair allocation. Here, B has a prescription value of $P_{C_B} = 1$ owing to her allocation power. A also has the same value, which again reflects the expectation not to delegate and to not allocate unfairly. Outcome values differ: B allocates unfairly, resulting in $O_{C_B} = 0$ and therefore $AD_{C_B} = 1$. A delegates but is not making

[18] The prescription values are arbitrary and can be assigned any number without loss of generality as long as an ordinal scale is maintained, allowing for differentiation of outcomes as fully, partially or not at all meeting the prescription.

[19] The non-weighted blame values are 0.8 and zero, respectively, but are weighted relative to the overall blame attribution. See definition 8 above.

an unfair allocation, therefore only failing to meet the non-delegation prescription, resulting in $O_{C_A} = 0.5$, $AD_{C_A} = 1 - \frac{0.5}{1} = 0.5$. We therefore predict individual blame values of 1/3 to player A and 2/3 to player B ($wb_{C_A} = 0.33, wb_{C_B} = 0.67$).

The predicted results are summarized in Figure 6:

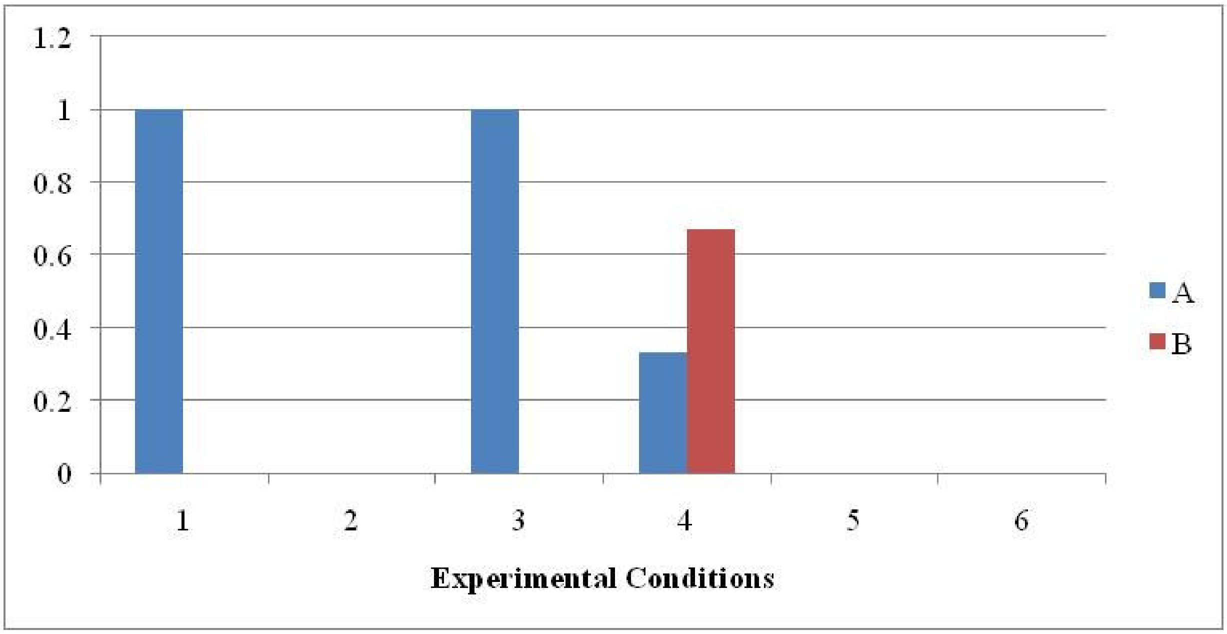

Figure 6 – predicted results

It can readily be seen that the model's predictions are in general agreement with the empirical results obtained by Bartling and Fischbacher (figure taken from Bartling and Fischbacher, 2008: 9):

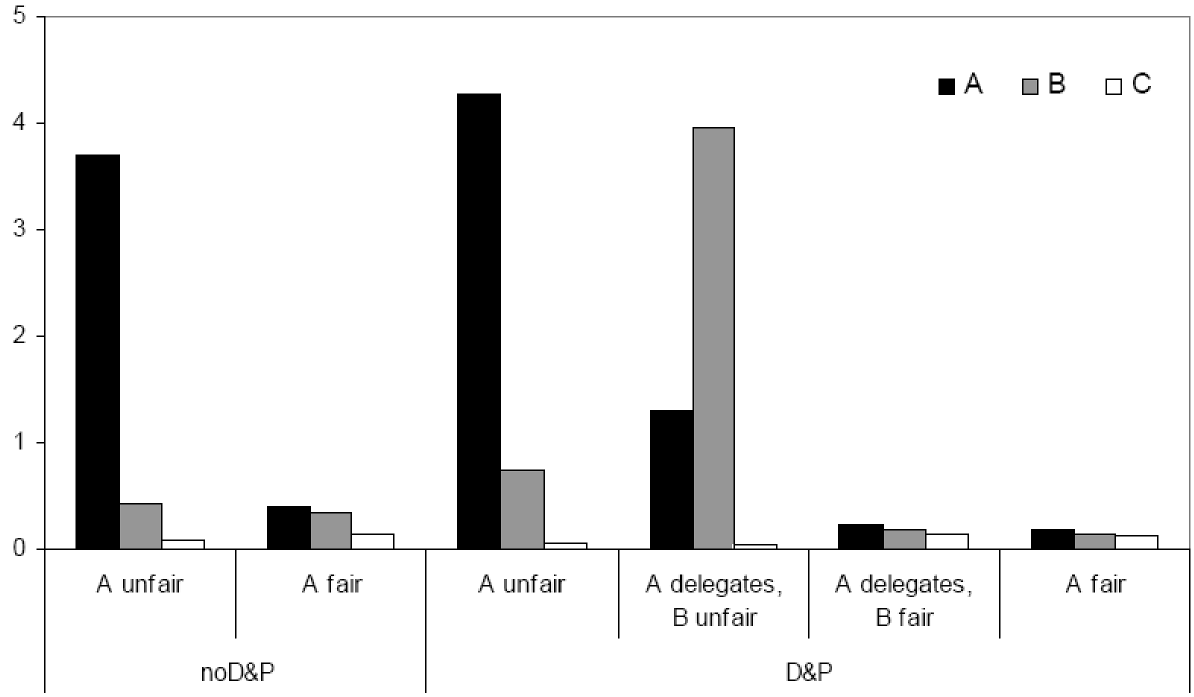

Figure 7 – empirical results from Bartling and Fischbacher (2008) experiment

Schlenker et al. (1994) discuss a prescription-identity-event triangle model of responsibility. As mentioned earlier, their experimental results support their hypothesis on the influence of structural links' strength on responsibility evaluation formation. Their results, while not directly implementable in our model[20], may give a prediction as to the relative importance of the different constructs. Specifically, they present findings that demonstrate which structural links were most sought after as sources of information for responsibility judgments by subjects. They determine that information on the identity-event and prescription-identity links were most important (selected by subjects 78-79% of the time in a responsibility-determining task), while the prescription-event link proved less pertinent (62%). Within the structure of our model, this may imply that both the agency dimension and the link between agency dimension and perceived loss are more important than perceived loss in determining the level of blame attribution. Schlenker et al. explicitly focus on responsibility, rather than blame, avoiding theorizing on the effects of perceived outcomes. Outcome bias has important implications on the shaping of judgments (Baron and Hershey, 1988), and so similar experimental designs that do incorporate scenarios where outcomes are either presented or not may provide a closer estimation of the true relationship between blame constructs.

Our model assumes that blame attributions made by individuals are on a cardinal scale: It is not only important to understand the ordinal scaling of agents made by each individual, but also it is essential to gauge the relative strength of blame attribution in order to attain a valid aggregate

[20] Both because of experimental design (and purposes), and the focus on responsibility rather then blame.

blame distribution that would represent what we postulate to be the public perception of blame. If, instead, an ordinal or dichotomous measure is used to evaluate blame attributions, we could still derive results through the model, but they would run the risk of being unrepresentative of the true dynamics and preferences of the population, and may actually display contradicting blame distributions.

This risk is demonstrated using data collected by Neil Malhotra and Alexander Kuo (2008) on public blame following Hurricane Katrina in 2005. Malhotra and Kuo asked surveyed subjects to rank, in an ordinal scale, seven prominent public officials (including President Bush and New Orleans Mayor Ray Nagin[21]), according to who should be blamed most for the loss of life and property damage caused by the storm. They presented results that rely both on the ordinal scale in its entirety, and on a second measure that focuses on the number of times the official was placed at the top of the blame scale, which is akin to an extended binary scale. Both measures are not in line with the processes that we believe control social blame formation, and produce contradicting blame distributions. To utilize this data set in the framework of our model, we translated both the original ordinal and quasi-binary (i.e. time-count) data attained from the 361 surveyed subjects into corresponding cardinal scales of 361 individual blame attributions. Figure 8 displays the resulting blame distributions produced by the model. The vertical axis is the value of each agent's weighted blame attribution:

[21] The public officials available for the respondents to rank were Louisiana Governor Kathleen Blanco (Democrat), Federal Emergency Management Agency Director Michael Brown (Republican), President George W. Bush (Republican), Secretary of Homeland Security Michael Chertoff (Republican), New Orleans Mayor Ray Nagin (Democrat), Louisiana Senator Mary Landrieu (Democrat), and Louisiana Senator David Vitter (Republican). The experiment itself was concerned with how information on political affiliation and public role affected blame attributions.

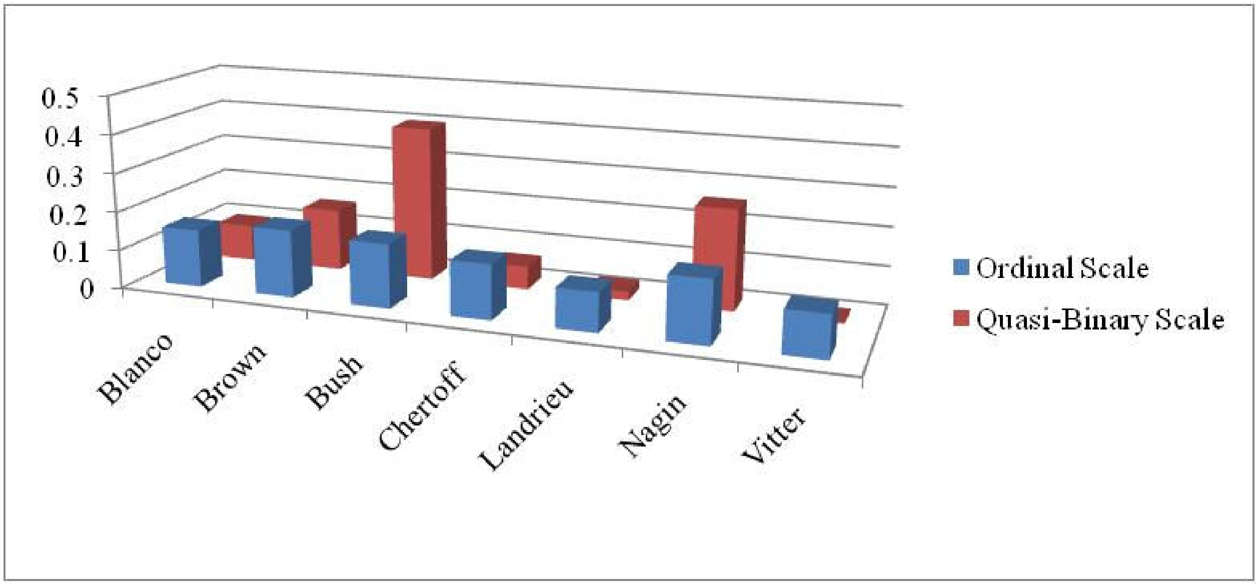

Figure 8 - model results for two different Malhorta and Kuo (2008) survey data sets

It is evident that the quasi-binary scale, which counts the number of times an official was ranked highest by subjects as deserving blame for Katrina, identifies President Bush as the main social-blame taker. In contrast, when taking into consideration the rest of the information contained in the ordinal scaling and using it as such, we attain a very different distribution where differences in blame attribution are flattened. Furthermore, the principal blame-taker becomes FEMA director Michael Brown, despite being ranked 3rd when using the quasi-binary scale. These results emphasize that in order to maintain consistency and validity using our model, interval data must be used.

Another debatable design choice is how we model prescriptions made by individuals. In our definition of prescriptions (both for events and agents) we make a conscious reduction of the wide scope of known features of agency and perceived loss attributions into a single, some would say meaningless, numerical value. If we were to try and provide a framework that expresses all of the different constructs of an individual's set of prescriptions on an issue we would have to come up with an extremely complex and highly contentious function that will most probably still be inadequate for many specific cases. Instead, we leave the prescriptive ingredients, their weights and the explicit prescription function to be determined on a case-per-case basis. This choice also allows for simplified designs that only focus on specific foci of attention of individuals in specific cases and experimental environments, without having to capture their entire set of considerations. An exemplary application of this idea is provided above in our simulation of the Bartling and Fischbacher (2008) experimental design. Of course, more complex cases can also be analyzed by introducing richer prescription functions, based on theoretical or empirical findings on the relevant prescriptions that are "at work" on a given issue.

4. Possible Effects of Social and Political Constructs on Social Blame

In the previous chapter we established a formal description of the formation of social blame. We now ask what social and political constructs may affect social blame creation and distribution, and how these effects are established. We focus on two concepts that may facilitate or constrain social blame attributions – institutional design and social cleavages[22]. In the subsequent chapter we proceed to develop measures for coherence and agreement within the framework of our model. These concepts and their interaction enable us to incorporate the assumed effects of social and political factors into the theoretical environment of social blame.

In the following short discussion of institutional design and social cleavages as variables that modify social blame, we can hardly begin to provide a serious review of these two copious subfields of political science. Indeed, we do not attempt to do so, but instead choose to discuss one theoretical aspect from each subfield that we hypothesize to be substantial and representative of the relationship between them and the development of social blame.

Political institutions are a central research subject in Political Science. They have been theorized and demonstrated to influence many aspects of political life, including, but far from limited to, voter behavior, legislation, international relations, the adoption of policies and democracy in general. We find it plausible to suggest that the institutional design of a society also has an effect

[22] This does not imply that other political concepts are impertinent to analyzing social blame. We in fact believe the opposite to be true, but in the limited scope of this work we only seek to demonstrate a possible connection between well-established political science concepts and the realm of blame. We therefore focus on two concepts in the hope that further, more extensive research will follow.

on the way social blame is shaped[23]. As one possible variable, we propose to look at the influence of a system's configuration of Veto Players (Tsebelis, 2002) on blame. The theory of veto players, as formulated by George Tsebelis, has been a prominent instrument in analyzing some of the fundamental questions that dominate the research on policymaking. The primal theoretical argument regarding Veto Players[24] is that their presence consolidates status quo and impedes political change through legislation. We hypothesize that the number of veto players in a system also constraints the public's ability to attribute blame clearly to a small number of agents. As more veto players are present, it becomes more difficult for individuals and for the public in general to identify a single agent as the sole blame-taker, and overall blame becomes incoherent. Conversely, when the system is characterized by few veto players, the public will tend to focus blame on the veto players that are present, thus creating a more coherent blame distribution. Our suggestion of a connection between the concept of veto players and blame is also in line with Alicke (200: 557, also see discussion in Chapter 2) emphasis on the importance of control in blame evaluation. The existence of many veto players reduces their causal control over the consequence of the decision, thus rendering the assessment of an agent's contribution to the harmful event more difficult, therefore producing weaker blame attributions.

It follows from our argument that institutional design (described and classified using the extant veto players framework) is expressed in agency blame coherence. We therefore complement this

[23] This is in line with various previous works that have demonstrated the effect of institutional structure (specifically, of divided government and high levels of coalition complexity) on issues like voter behavior and economic voting (Anderson 1995, 2000; Lewis-Beck 1988; Leyden and Borrelli 1995; Powell and Whitten 1993; Rudolph 2003a).

[24] Veto Players are individual or collective actors that have to give agreement in order for a policy to change (Tsebelis, 2002:2).

discussion later by formulating a measurement of agency coherence that functions as a formal expression of the theorized effects of institutional design on blame structure.

Similar to institutional structures, social cleavages also have profound effects on a wide range political science issues and concepts. The number and type of social cleavages have been shown to interact with, among others, the stability of political systems, voting patterns, and then number of parties (see for example Neto and Cox, 1997, who also provide an overview of the writing on social cleavages). In fact, institutional structure and social cleavages have long been competing explanations for the structure of party systems (Mozaffar, Scarrit and Galaich, 2003: 380). Here, too, we suggest that social cleavage is one of the factors affecting social blame distribution and formation. We assert that the nature of a society's fragmentation serves as a constraining variable that, together with other factors, determines the nature of how the public allocates blame, and particularly how homogenous or heterogeneous the social blame map is – namely, the level of agreement within the public over who to blame. The saliency of social cleavages as determinants of social blame depends upon an interaction with the type of issue at hand. A society may be deeply divided over one public issue (e.g. economic ideology), but if the issue at hand is orthogonal to opinions on this divided dimension (e.g. a failed response to a natural disaster), it will not amplify public disagreement. In contrast, when the issue is related to the dimension(s) over which the society is fragmented, we expect overall blame patterns to be in stronger disagreement, as different sections of the public will tend to blame different agents for their perceived loss, in correlation with their differing positions and biases towards agents. To accommodate this effect within the framework of our model we therefore develop a measure of agreement that can give rise to social cleavage-based analysis of aggregate blame.

The following formulas summarize the hypothesized effects excreted on blame by institutional design and social cleavages:

Institutional design:

$$\uparrow \text{Number of Veto Players} \quad \Longrightarrow \quad \downarrow \text{Blame coherence}$$[25]

Social structure of society:

$$(\uparrow \text{Social cleavages}) \times \text{Issue-related cleavage} \quad \Longrightarrow \quad \downarrow \text{Blame agreement}$$[26]

Finally, note that according to the veto players theory, an increase in the number of veto players makes policy change harder to attain, while a small number eases deviations from the status quo. Assume further that our argument on the connection between the number of veto players and blame coherence also holds. It therefore follows that correlation between policy change and blame distribution is to be expected, as both are (at least partially) governed by the number of veto players in the system. Specifically, when policy is prone to change, we can expect to see more coherent blame distributions regarding the measures being suggested. In contrast, when status quo is stable, blame distribution will tend to be incoherent.

[25] As referred to above and as discussed in Chapter 1. For an explicit definition of our measure of the concepts of coherence and agreement, see Chapter 5.

[26] See previous footnote.

5. Coherence and Agreement

In chapter 2 we discussed the theoretical concepts of coherence and agreement. We now continue with the formal definition of these concepts within the framework of our model, with the final goal of achieving a conceptual framework that allows for the introduction of social and political variables, presented in chapter 4, to the analysis of blame.

We first develop a measure of agency coherence, and then introduce a measure for the concept of agreement. Next, we analyze the interaction between these two concepts and describe social and political situations that characterize blame scenarios that combine different levels of coherence and agreement.

Agency Coherence

Measuring the concept of coherence requires better understanding of when a specific blame distribution can be considered to be coherent. In the following Figure 9, the bars represent two possible distributions of blame over two agents:

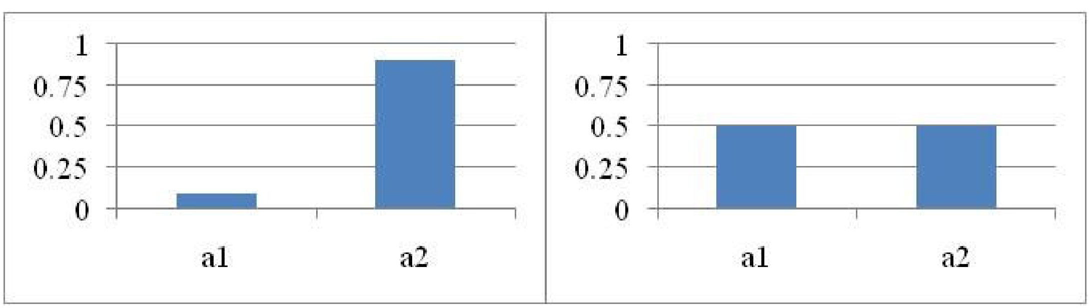

Figure 9 - two possible blame distributions over two agents

When the aggregate blame dispersion on an issue is extremely contrastive – i.e., directed almost entirely at one agent, as demonstrated by the left plot $\left(wB_{a_1} = 0.1, wB_{a_2} = 0.9\right)$, we can determine that the aggregate blame dispersion is highly coherent. In the right plot, aggregate blame is equally distributed over the two agents involved on an issue $\left(wB_{a_1} = wB_{a_2} = 0.5\right)$, which labels the situation incoherent, as neither agent can be singled out as the central blame-receiving figure. These cases, however, represent the extreme ends of the scale of distributions. Consider the following blame distribution:

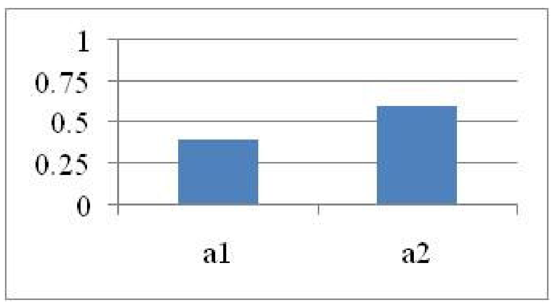

Figure 10

While it is evident that a_2 is attributed with a higher value of aggregate blame than a_1 ($wB_{a_2} = 0.6$ while $wB_{a_1} = 0.4$), similar to the Figure 9, the level of coherence in this scenario is lower, as it can be inferred that public opinion, although inclined to see a_2 as the main blame carrier, is not as decisive as in the first scenario.

Other possible instances entail further challenges to defining the concept of coherence. Consider the following two scenarios depicted in Figure 11:

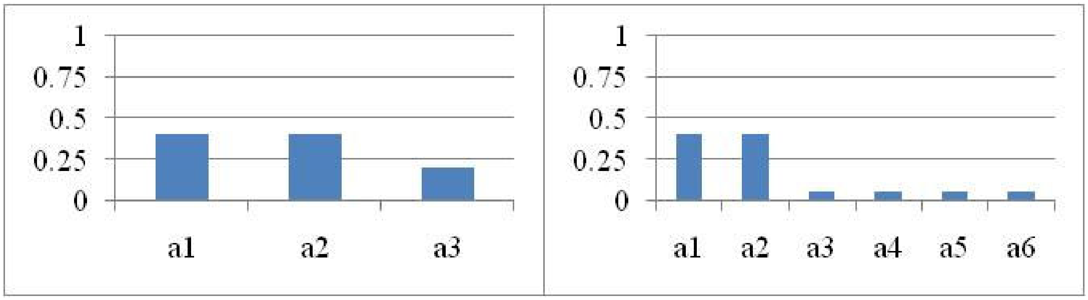

Figure 11 - similar blame distributions over a varying number of agents

In both cases, $wB_{a_1} = wB_{a_2} = 0.4$. Yet, the first scenario consists of three agents and the issue depicted in the second involves six agents. In our opinion, this has an important effect on the aggregate blame coherence, and we postulate that coherence is higher in the second scenario: In the first issue depicted, we predict that most people will consider the situation to be confusing or at least unclear ("It is hard to decide who the main blame-taker is"). In contrast, the second case is more likely to produce a coherent impression. We believe people will be more inclined to asses this distribution as one that portrays clear blame attributions to a_1 and a_2 over the rest of the agents, rather than be confused by which of a_1 and a_2 is the main blame-carrying figure.

Measuring coherence is conceptually similar to the use of centrality measures of distributions, which are used as indicators in a range of scientific fields. Common examples include the Concentration Ratio (Cowling and Waterson, 1976: 271) the Herfindahl – Hirschmann Index (or HHI, see Hirschmann, 1964), used in economics to measure competitiveness; indices of party classification in political science, like the Taagepera and the Kesselman-Wildgen measures (Molinar, 1991); the Gini coefficient, Atkinson indices and other measures of inequality (Amiel and Cowell, 1999); a large number of measures of ecological diversity (for a comprehensive review see Magurran, 1988); different entropies and divergence distances (especially the Shannon Entropy and its derivatives, like the Kullback-Leiber divergence (Kullback, 1983, Lin, 1991)); and the kurtosis level of probability distributions. A distribution's variance can also be used to indicate how concentrated is the data relative to the mean.

To choose the appropriate measure to be used for evaluating coherence, we conduct a comparison of seven indices'[27] performance when applied to the measurement of agency coherence (see appendix for a detailed comparison). Going through three different sets of comparisons, each incorporating scenarios with varying degrees of coherence as we conceptualize it, we gradually eliminate candidate measures and are left with the preferred measure. We first compare five different four-agent distributions with increasing levels of blame concentration. We then perform a similar comparison using eight 10-agent distributions.

The indices that displayed measures consistent with our specifications were then committed to a third set of distributions where in all cases only one agent receives all of the blame, but the number of overall agents varies. Expecting to see a growth in coherence as agent numbers increase, we are left with a single measure of coherence based on the KL entropy that satisfies this condition. Seeing that KL entropy is the only measure consistent with our conceptualization of coherence, we therefore choose a measure of aggregate agency coherence based on it.

Originally, KL entropy is used to compare two distributions and determine whether a sample which is hypothesized to belong to a certain distribution does resemble it with a predetermined confidence level. In our adaptation of the formula, we compare the aggregate blame distribution

---

[27] Variance, Herfindahl-Hirschman (HHI), Normalized HHI, Index of Dispersion, Kullback-Leiber divergrence (KL entropy), Gini coefficient, Kesselman-Wildgen Index (KW). Other surveyed indices were eliminated for being mathematically equivalent to one or more of these indices, and for relying on information irrelevant or missing from our model (especially the level of kurtosis, which, although computable, is inconsistent and lacks explanatory power because of not being applied to a single probability distribution).

to a uniform distribution over all the agents, hereby represented by N samples, each with a size of $1/_{|A^s|}$. In essence, then, we measure the blame distribution's deviation from a uniform distribution of blame that represents complete incoherence. This reasoning is expressed in the following formulation:

Definition 12: Let $A^s = \{a_1 \cdots a_N\}$ be the set of agents on an issue *s*. The agency dimension *coherence* on an issue *s* is defined as:

$$C^s = \sum_{k=1}^{N} wB_{a_k} \ln\left(\frac{wB_{a_k}}{1/_{|A^s|}}\right)$$

Where wB_{a_k} is the weighted aggregate blame for agent a_k. While C^s has a lower bound of zero[28], its upper bound is dependent on the number of agents involved on an issue. For example, when $|A^s| = 2$, $C^s \leq 0.69$, whereas if $|A^s| = 4$, $C^s \leq 1.38$, and if $|A^s| = 10$, $C^s \leq 2.3$. This entails the view that higher coherence values are only attainable when agents stand out of a background of a relatively large number of agents[29]. (insert graph of 1-100 max coherence values, include different coherence scenarios) Figure 12 displays the maximum attainable coherence values as a function of the number of agents present on an issue. In all instances, the maximum value is attained when blame is directed solely towards a single agent.

---

[28] This value can only be attained if all agents receive equal blame ($wB_{a_k} = wB_{a_l}\ \forall a_k, a_l \in A^s$).

[29] We do, however, assume that on most issues, large numbers of blame-carrying agents are highly improbable. Consistent with this, and since even when $|A^s| = 100$, $C^s \leq 4.6$, we can set an arbitrary bound of C^s at the value of 5.

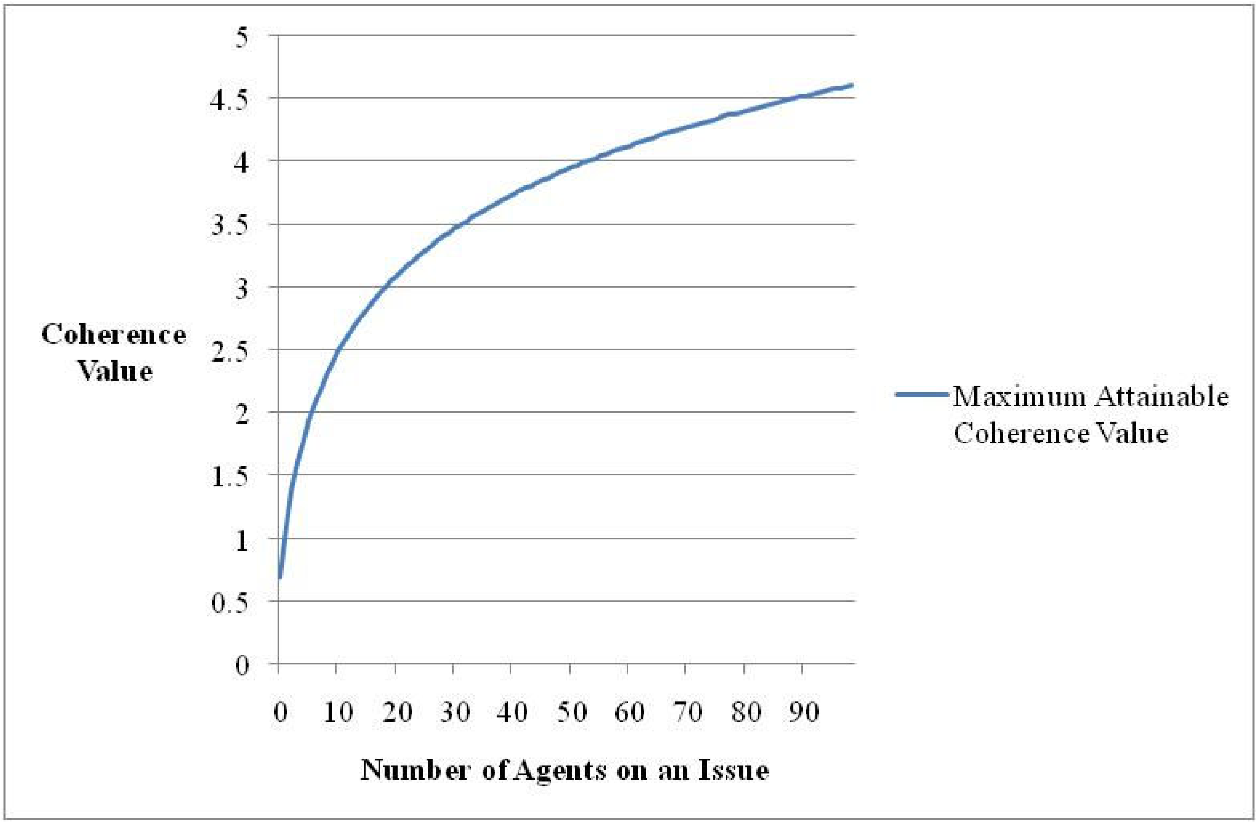

Figure 12 – maximum attainable coherence values with varying number of agents

Agreement

Different aggregate blame distributions (and their accompanying coherence values) are the result of the amalgamation of blame attributions made by a large number of individuals. It is therefore possible for a single blame distribution to be the product of several different opinion distributions within a population. In this respect, the concept of coherence cannot, on its own, be a sufficient measure of how social blame is assigned: A very coherent blame distribution may turn out to represent disagreement between several subgroups or sectors of the population, while an incoherent blame map can still reflect unanimity of opinions (i.e. "everyone is equally confused"). Consider the following aggregate blame distribution over two agents, a_1 and a_2, which is derived from a population of 100 individuals:

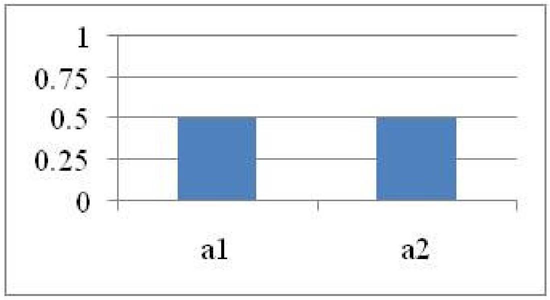

Figure 13

While this distribution is highly incoherent ($C^s = 0$), it can represent various blame distributions among the population, and in particular, two very contrasting schemes. In the first scheme, every individual has a blame attribution of 0.5 for each of the two agents ($\forall i, wb_{i_{a_1}} = wb_{i_{a_2}} = 0.5$). In the second, half of the individuals strongly blame a_1 and carry no attribution towards a_2 - their weighted individual blame attributions are [1,0], while the other half strongly blame a_2 and feel that a_1 is exculpable, and so have a vector of [0,1]. Both these schemes produce the same aggregate blame attribution. Yet, the first reflects full agreement between every two individuals on how blame is to be assigned to each agent. In contrast, the second scheme portrays strong disagreement over who is to blame between two mutually exclusive and exhaustive groups of the population – a disagreement that results in indecisiveness (= incoherence) at the public level . An essentially equivalent case of similar coherence values with contrasting agreement schemes appears in McGraw (1991: 1142)[30].

This concept of *agreement* within a population on how blame is attributed complements the notion of coherence. Coherence is a descriptive measure of how concentrated the aggregate

[30] Specifically, this is found when making a comparison of blamelessness between the two cases of “plea of ignorance” and “benefits”, who on aggregate display similar blame levels, but high agreement (i.e. low variation on account satisfaction) on the former and low agreement on the latter.

blame is over a small number of agents. By implying a measure of agreement we gain further insight into how representative this distribution is of public opinion – whether it reflects a divided society or a consensus among its members.

An attempt to devise a measure of agreement faces several challenges. One starting point can be to measure the coherence of each individual's weighted blame vector[31], and use it as an indicator of each individual's personal blame distribution. With every individual having this value assigned, we can now look at their mutual agreement by calculating the variance of coherence values over all individuals in the population. This is actually a measure of *dis*agreement, as a value of zero represents complete unanimity among the individuals' coherence values. Although intuitive, the calculation has a major pitfall: When applying it to the two examples provided above, they both yield full agreement (a value of zero), although the second case (half of the population blame a_1 while the other half blame a_2) is highly disagreeable[32].

Alternatively, to be able to discriminate between these two schemes that result in the same (incoherent) distribution of [0.5,0.5], we can compute the aggregate variance over each agent, and take their mean. In the first case, where $\forall i, wb_{i_{a_1}} = wb_{i_{a_2}} = 0.5$, we get zero variance for a_1, and zero variance for a_2 (as each agent assigns the same value for both agents), and therefore

[31] For example, if on a given issue there are four agents involved, then the individual has a weighted blame vector of size four. If for a given individual this blame vector is: [0.1,0.2,0.2,0.5] then this individual's coherence is 0.1657, using the previously proposed measure of coherence.

[32] In the first case, every individual will have the same (zero) coherence, and therefore a variance value of zero. In the second case, while each half of the population has contrasting blame attributions, they are essentially inverse and the coherence value for an individual from one half will be the same as for an individual from the other half. This too leads to zero variance of coherence values.

a mean of zero. In the second case the outcome is different: a_1 receives an individual blame value of 1 from 50 individuals, and a value of 0 from the other 50, which results in variance of 0.25. The same applies for a_2 and therefore the mean is also 0.25, allowing us to infer that it represents public disagreement over the aggregate blame – while the first case represents agreement within the population over the result.

This heuristic encounters problems, too. Consider the following aggregate blame distributions: The first is identical to the example discussed above: two agents, a_1 and a_2, with equal blame values: $wB_{a_1} = wB_{a_2} = 0.5$. The second consists of ten agents $a_1 \cdots a_{10}$, with $wB_{a_1} = wB_{a_2} = 0.46$ and $wB_{a_3} = \cdots = wB_{a_{10}} = 0.01$ (i.e., the first two agents share most of the blame while the others carry very small blame attributions). This is demonstrated in Figure 14:

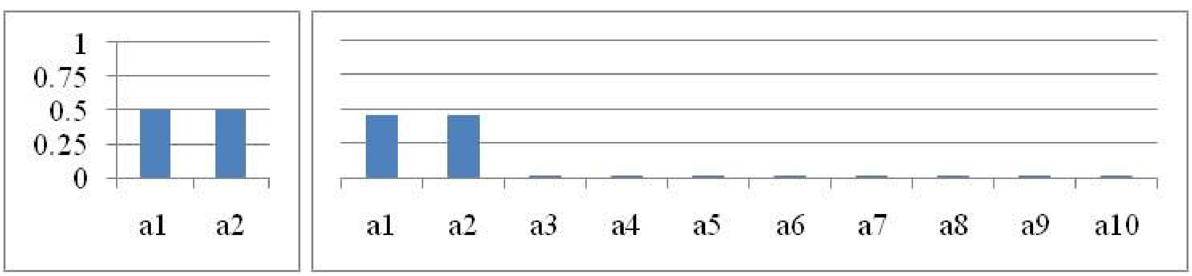

Figure 14 - two similar blame distributions varying on number of agents

Assume that in the first case, the population is divided as in the second scheme from the last example – 50 individuals with blame vectors of [0,1] and 50 with [1,0]. We assume a similar division in the second case as well: 50 individuals with [0.91,0.01,0.01…0.01], and the other 50 with [0.01,0.91,0,01,…0.01]. Using the aforementioned calculation, we obtain an agreement value of 0.25 for the first case, and 0.0405 for the second. Yet, it is difficult to accept the resulting interpretation that individuals in the second case are in a stronger agreement with each other than those in the first (more than four times smaller in value), since they only agree on who *not* to blame.

We therefore introduce a measure of agreement that consists of summing agency blame variance values. This measure is indifferent to increases in the number of agents when blame remains concentrated on the same agents, thus overcoming the problem presented above[33].

Furthermore, we would like the measure to provide a distinction between disagreements among few groups in the population (e.g. half the population blames one agent, and the other half blames another) and disagreements where many fractions of the population respectively blame different agents (e.g. ten sectors of the public blame ten different officials). We conceptualize the latter scenario, or even more fragmented cases, as representing stronger disagreement than when the population is split between only two groups. Summing over variances provides results consistent with this view.

Definition 13: Let $A^s = \{a_1 \cdots a_N\}$ be the set of agents on an issue *s* and $I^s = \{i_1 \cdots i_M\}$ the set of individuals on the issue. The population *agreement* on an issue *s* is defined as:

$$AG^s = \sum_{k=1}^{N} \sigma^2 \left(wb_{i_{1_{a_k}}} \cdots wb_{i_{M_{a_k}}} \right)$$

Where $\sigma^2 \left(wb_{i_{1_{a_k}}} \cdots wb_{i_{M_{a_k}}} \right)$ is the variance of all the weighted individual blame attributions directed at a_k.

[33] If an agent is added to the issue but attracts little blame attribution, her blame variance remains low (or zero, if no blame at all is directed at her). When summing for variances, this addition will have either marginal or no effect at all, thus preserving the same magnitude of agreement.

Using this definition, in the example described above we obtain agreement values of 0.5 in the two agent case and 0.405 in the ten-agent case (with $|I^s| = 100$), indicating strong disagreement in both scenarios, while still differentiating between the different blame distributions that characterize the populations in each case.

AG^s has a lower bound of zero, signifying complete agreement between all individuals over how blame is distributed over the participating agents. The absolute upper bound is 1, theoretically attainable only when the number of agents is infinite, and lowers as fewer agents are present: a two-agent issue has a maximum AG^s value of 0.5, three agents can reach a maximum of 0.66 and the ten-agent scenario has an upper bound of 0.9. This design therefore entails the idea described above that the presence of more agents creates the potential for stronger disagreement within a public. The attainment of these maximum values is dependent on how fragmented the public's blame attribution is[34]. Figure 15 displays the maximum attainable agreement values (meaning the most disagreement) as a function of the number of agents present on an issue:

[34] Maximum disagreement is attained when a public is proportionally divided between all present agents. For example, when there are two agents, maximum disagreement is possible when half the public blames one agent, while the other half blames the second. When there are three agents present, we achieve maximum disagreement when a third of the public solely blames the first agent, another third blames the second agent, and the last third blames the third agent. Similarly, when there are ten agents present, maximum disagreement is attained when each agent is blamed by a tenth of the population, and only by that tenth.

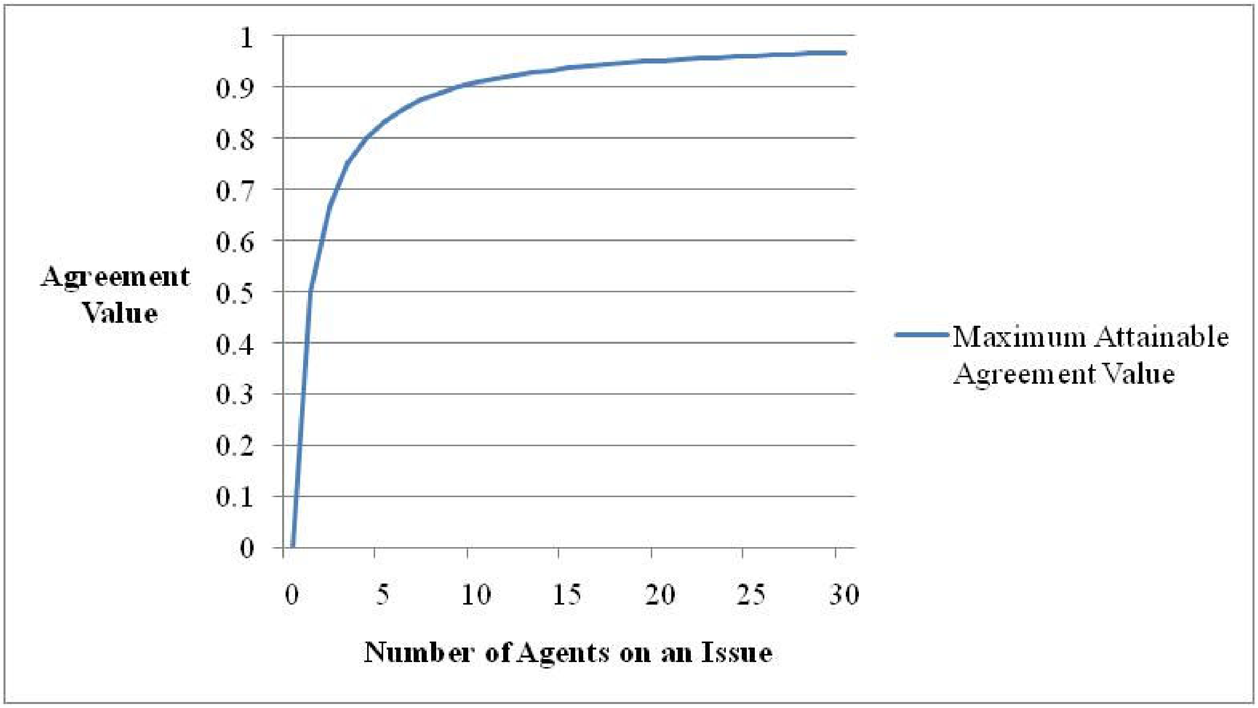

Figure 15 - maximum attainable agreement values with varying number of agents

Interactions between Agency Coherence and Agreement

We can now utilize the concepts of coherence and agreement to classify social blame situations and attribute them to different states of institutional design and social cleavages.

As we have already seen when developing the concepts of coherence and agreement, it is important to note that there is no a-priori correlation between them: an issue may exhibit an incoherent blame dispersion that was obtained by either full agreement of all the individuals involved, or by a disagreeing crowd. There is, however, some level of dependence between the two measures that becomes apparent as we compare the possible pairings of coherence and agreement measures.

High Coherence / High Agreement:

Instances where the agreement within the population is high and the aggregate blame perception is coherent are mostly events in which a small number of agents receive most or all of the blame and there is strong consensus regarding their identity. A typical case would be when a politician is found guilty of corruption, and so is a single target of blame attributions by all or most of the population, or where there is a large scale scandal related to a very clear hierarchical structure and institutional responsibility expectations. An example is the UK rail network paralysis following several crashes in 2000, where more than half of the surveyed population held the infrastructure provider Railtrack to be greatly to blame, and in contrast attributed significantly smaller blame values to the government, thus displaying high agreement on a coherent blame attribution (Hood, 2002).

In terms of institutional design and social cleavage, we postulate that an institutional setting with a small number of veto players, coupled with low social divide (or an issue which is independent of social cleavages) will lead to a high coherence - high agreement aggregate blame distribution.

Although high coherence can be attained with more than one agent being the principle blame-carrier, the high level of agreement constrains the possible number of blame-taking agents: If there is high agreement that many agents are to be blamed on an issue, this will inevitably be expressed in low coherence values. From this it follows that high coherence - high agreement arrangements involve a small number of blame-takers.

High Coherence / Low Agreement

Issues that couple high levels of coherence together with public disagreement can represent a distribution of blame over a small number of agents (but more than one), where several sectors of the population each see one of these agents as the sole blame-taker. This, we assert, is characteristic of a polarized society where contrasting accounts are adopted by rival factions or partisan camps. Specifically, where power is shared by a two or three-party coalition government, we may witness supporters of each party blaming the other party's representatives for an unpopular decision made by the government, thus focusing blame on two or three agents while maintaining low overall agreement. This configuration is akin to the case where veto players vary in partisan affiliation, especially within coalitions (i.e. Partisan Veto Players, see Tsebelis 2002).

Veto Players theory argues that the more fragmented coalitions are, the more veto players will be present (Tsebelis, 2002: 93-99). Our implementation of the theory of veto players to the concept of social blame coherence suggests that with an increase in the number of veto players, the probability of attaining high coherence will decline. Therefore, high coherence with low agreement will most likely be induced by a small number of partisan veto players, a situation that enables strong blame concentration but involves noticeable social divide.

Furthermore, low agreement is theorized to be derived from a social cleavage pertinent to the issue at hand. In this case, the high coherence condition bars this divide from being characterized by many sub-groups of the population who hold different opinions. Instead, it is most probably a result of one major public opinion fault line (e.g. Democrats-Republicans, or Liberals-Conservatives) where each group holds different veto powers (for example, a Democratic President with a Republican-controlled Congress in the United States).

Other instances that lead to high coherence – low agreement are also possible. Consider a situation that involves ten agents, and a population of 300 individuals divided into three sectors, with each sector only blaming a single agent (a_1, a_2 and a_3, respectively). This results in the following aggregate blame dispersion[35]:

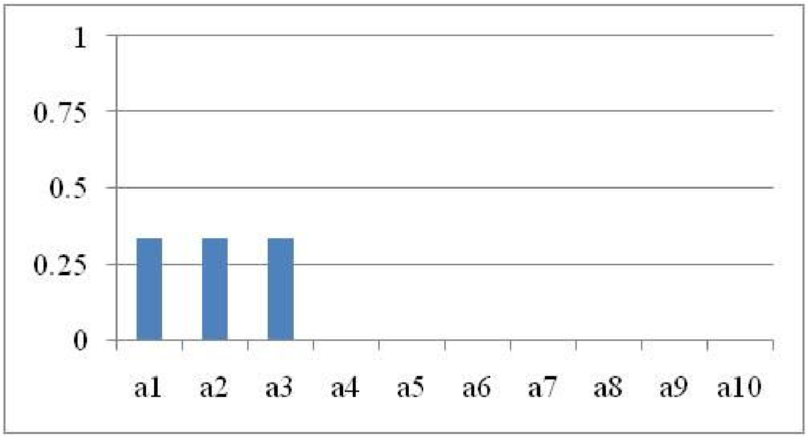

Figure 16 - example of high coherence – low agreement distribution

This scenario carries a coherence value of 1.203, which is high considering that the upper bound for the coherence value on an issue with $|A^s| = 10$ is 2.3. In contrast, the agreement value is 0.666, representing strong disagreement.

High coherence with low agreement can also be obtained in instances where there is disagreement between sub-groups of the population over several blame-taking agents, yet one of the agents receives blame from more than one sub-group, and so in aggregate, turns out to carry strong blame in a coherent manner. If, for example, a population of 300 individuals is divided as before between three sectors, and each sector mainly blames either a_1, a_2 or a_3, respectively, but also directs some blame at a_4[36], the resulting distribution makes blame concentrate on a_4:

[35] Similar to a three-party coalition government of contrasting ideologies, as is common in Israel in recent years.

[36] An individual from the first sector would have a blame vector of [0.7,0,0,0.3,0….0]. An individual from the second sector would have a vector of [0,0.7,0,0,0.3,0…0], and so a4 receives a weighted blame value of 0.3 from

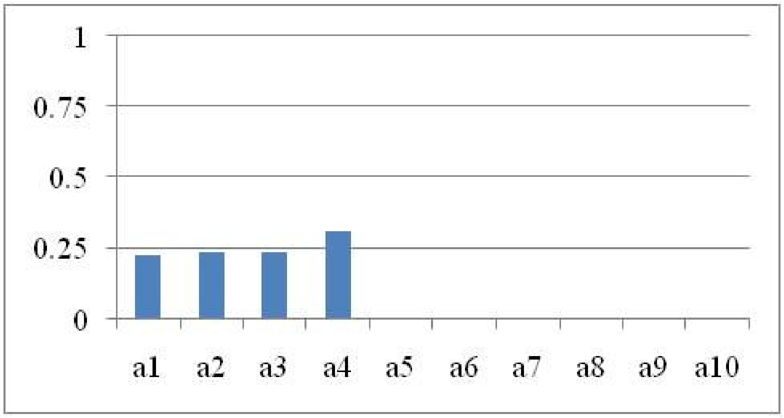

Figure 17 - second example of high coherence – low agreement distribution

Coherence remains high in this scenario, with a value of 0.924, and there is notable disagreement measuring 0.3266, although less than in the previous example.

A high-coherence low-agreement scheme was induced by Arceneaux and Stein (2006: 45-48), who asked respondents to indicate what level of government should get the blame for the city of Houston, Texas being unprepared for the 2001 flooding caused by Tropical Storm Allison. Surveyed subjects were asked to name one of either the national, state, county or city level as the blame-taker. This design produced the following weighted blame distribution over the different government levels, based on the survey results:

every individual in the population, while each one of a1,a2,a3 receive a weighted blame vector of 0.7 from a third of the population.

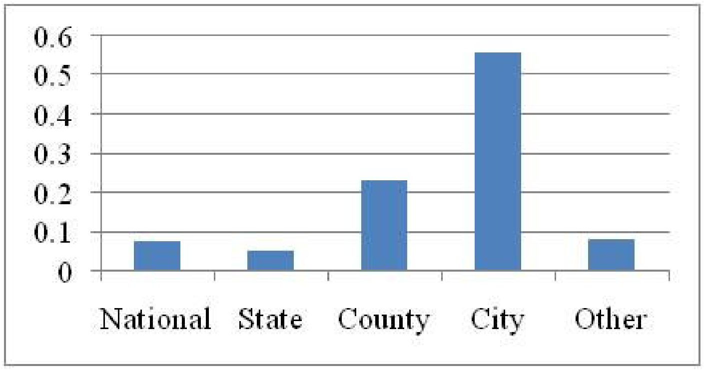

Figure 18 - model results for Arceneaux and Stein (2006) data

Although carrying high coherence, consistent with the strong concentration of blame on the Houston municipality, the survey design results in an agreement value of 0.687 (with maximum attainable disagreement being 0.83), indicating strong disagreement within respondents over the results. However, this is attributed mostly to the question format. If, in contrast to what has been carried out, the respondents would have been asked to assign blame values to each of the government levels instead of pointing one out as the sole blame-taker, less disagreement would most likely be attained.

Low Coherence / High Agreement

High agreement on a low coherence distribution of blame means that there is public consensus that no one agent or even a small group of agents carry most of the blame on an issue. Where coherence is low we expect to encounter a more or less equal blame distribution over a large portion of the agents on an issue. High agreement means that this distribution of blame is shared by most of the individuals. Therefore, the public is inevitably accordant on spreading the aggregate blame over a large portion of the agents on an issue, which represents a prevalent inability to point out one figure which is portrayed as the sole blame-carrier.

Institutionally, this distribution could be obtained when there are many veto players in the system, reducing the ability to draw clear blame attributions towards a small number of agents. In this analysis, the high agreement condition implies that either the issue is unaffected by social cleavages, or that society in general is not divided or polarized.

With regards to issue content, we postulate that situations of this kind involve cases where the human effect is considered to be only secondary to other factors, particularly where damage is inflicted in a manner that is considered to be unavoidable or unpredictable by the 'usual suspects'. This may be characteristic of some economics issues. Events like a widespread economic crisis, stock market crash or a state-wide union strike may be perceived similarly to Force Majeure occurrences in that no particular agent/s can be held accountable or responsible for precipitating the outcome, while large parts of the population are subjected to extensive loss, therefore resulting in similar coherence and agreement patterns.

A striking example is provided by Debra Javeline (2003), who analyzes survey data on protest participation in Russia following a massive wage crisis in 1998. Javeline finds that the propensity to demonstrate is strongly correlated with the specificity of blame attributions carried by citizens, and demonstrates that the large majority of the population (87%) held vague and unspecific blame attributions, thus being unable to derive causal attributions to a small number of specific agents, and instead either attacked a large number of public figures, or blamed more abstract concepts like "the central authorities" or "the general economic situation" (pp. 111-112).

While overall blame was highly incoherent, no particularly contradicting patterns of public blame were found, indicating that the blame image carried high overall agreement.

Low Coherence / Low Agreement

An issue that involves disagreement and incoherence will typically be the outcome of a divided population (i.e. having significant social cleavage). Where the population's blame attributions vary over a small number of groups, it is expected that neither group will, on its own, display a coherent (and distinct) aggregate blame distribution, as their overall aggregation will also result in a coherent distribution. The more fragmented the population is with respect to the various agency blame attributions (i.e. smaller groups with contrasting blame attributions), the more coherent each sub-group's distribution of blame can be while still retaining the aggregate result of low coherence.

We assert that these settings are relatively uncommon and are characterized by instability over time and issue formation. The behavioral dynamics of blame validation (Alicke, 2000: 558) and the tendency to seek the human causal component of events (ibid) governs individual's proclivity to converge unto an explanation of an event that underscores the agency factor, therefore making people uncomfortable without consolidating a narrative that allows them to denote human actors as responsible for their perceived loss (Felstiner et al., 1980, Schlenker et al., 1994: 634). Low coherence and low agreement imply that the situation is not only 'confusing' in general (i.e. incoherent), but also that people tend to feel differently about how exactly this confusion is outlined. We believe this is fertile grounds for opinion-shaping behavior by politicians, and interaction and deliberation between sub-groups of the population, and that under these conditions, and provided the issue remains salient, the individual and aggregate blame distributions will over time tend to converge to one of the three previously described pairings of coherence and agreement levels.

The suggested typology of social blame issues according to levels of agreement and agency coherence is summarized in Table 2:

Coherence / **Agreement**	*High*	*Low*
High	Few, probably one, blame-takers. *Typical scenarios*: Corruption accusation, clear hierarchy and expectations. *Institutional/Social makeup*: Few veto players, social cleavage orthogonal to issue	Public confusion over who is to blame. *Typical scenarios*: economic/financial crisis (Javeline, 2003). *Institutional/Social makeup*: Many veto players, homogeneous public opinion with regards to the issue
Low	Few (but more than one) blame-takers; polarized society. *Institutional/Social makeup*: Coalition government / strong partisanship (partisan veto players). 'Fault line' social cleavage.	Fractioned and undecided public opinion. Unstable situation that will tend to converge into one of the other three cases.

Table 2 – summary of suggested coherence and agreement typology of social blame

6. Summary and Future Development of the Model

This thesis contributes to the current body of research on social blame in three distinct ways:

First, the formal model of aggregate blame presented here integrates and synthesizes extant theories of social blame formation, drawing on common psychological and social elements that appear in the literature, while unifying overlapping and contrasting concepts. In this respect, we offer a more consistent conceptualization to be used by scholars interested in social blame research.

Second, our model incorporates an agency-based, perceived-loss dependent design, thus incorporating together micro-level individual blame attributions with the formation of macro-level social blame. This architecture allows for psychologically-grounded analysis of agency perceptions of blame. It also lends itself to experimental examination and implementation on existing data, from the fields of political science and psychology, as demonstrated in the text.

Third, we introduce the concepts of coherence and agreement of social blame, allowing for a more profound discussion of public blame formation. These concepts also pave the way for assertions on a proposed causal link between institutional design and social cleavages, and blame formations – connections yet unexplored, that can be attended to within the framework of our model.

Nevertheless, this work serves primarily as an introduction to what we conceive to be a promising horizon of future research of the attributes, dynamics and underlying variables that produce social blame. We believe this work lays solid foundations and opens the door for progress on of the following subjects:

The initial motivation that triggered this work was to provide a model that will describe both how blame avoidance motivation is initiated among politicians and public officials, and how their subsequent blame avoidance behaviors affect the blaming public, in accordance with the classification offered by Sulitzeanu-Kenan and Hood (2010). The design chosen for this model indeed meets these prerequisites. With relatively straightforward extensions to the static model presented in this thesis, we can transform it into a dynamic, time-framed model that incorporates agents' response to blame and captures the different blame avoidance strategies and behaviors that were classified in the literature. However, this endeavor is beyond the scope of this thesis.

A further benefit in progressing towards a dynamic model of blame and blame-avoidance behavior is the opportunity to further introduce cognitive considerations into the model's design. A dynamic design that tracks individual behavior over time allows for the incorporation of insights from cognitive psychology, especially those that involve attention constraints on complex causality. These additions can help create a richer behavioral basis for the formation of blame attributions, possibly allowing for the testing of hypotheses regarding blame convergence over time.

Lastly, this work attempted to connect social and political concepts dominant in the literature on politics with the formation of social blame. Yet our exploration of this research avenue is extremely limited in scope, owing to the nature of this paper. Nevertheless, we strongly encourage an examination of the influences that social and political design may have on a population's propensity to blame officials, and the range of possible blame distributions created in this process. We believe that identifying and describing the nature of these connections can expand what we see as a still incomplete explanatory framework of social blame.

7. References

Alicke, Mark D. 2000. Culpable Control and the Psychology of Blame, *Psychological Bulletin* 126(4): 556-574

Amiel, Yoram and Frank Cowell. 1999. *Thinking About Inequality*, Cambridge: Cambridge University Press

Anderson, Christopher J. 1995. *Blaming the Government: Citizens and the Economy in Five European Democracies*, Armonk, NY: M. E. Sharpe

Anderson, Christopher J. 2000. Economic Voting and Political Context: A Comparative Perspective, *Electoral Studies* 19: 151-170

Arceneaux, Kevin. 2003. The Conditional Impact of Blame Attribution on the Relationship between Economic Adversity and Turnout. *Political Research Quarterly* 56(1): 63-71

Arceneaux, Keving and Robert M. Stein. 2006. Who is Held Responsible When Disaster Strikes? The Attribution of Responsibility for a Natural Disaster in an Urban Election, *Journal of Urban Affairs* 28(1): 43-53

Arnold, R. Douglas. 1990. *The Logic of Congressional Action*, New Haven: Yale University Press

Baron, Jonathan and John C. Hershey. 1988. Outcome Bias in Decision Evaluation, *Journal of Personality and Social Psychology* 54: 569-579

Bartling, Bjorn, and Urs Fischbacher. 2008. Shifting the Blame: On Delegation and Responsibility, (http://ssrn.com/abstract=1166544)

Bovens, Mark, Paul 't Hart, Sander Dekker, and Gerdien Verheuvel. 1999. The Politics of Blame Avoidance: defensive tactics in a Dutch crime-fighting fiasco. In *When Things go Wrong*, edited by H. Anheier. Thousand Oaks: Sage

Brändström, Annika, and Sanneke Kuipers. 2003. From 'Normal Incidents' to Political Crises: Understanding the Selective Politicization of Policy Failures, *Government and Opposition* 38(3): 279-305

Burstein, Paul. 2003. The Impact of Public Opinion on Public Policy: A Review and an Agenda, *Political Research Quarterly* 56: 29

Cowling, Keith and Michael Waterson. 1976. Price-Cost Margins and Market Structure, *Economica* 43: 267-274

Ellis, Richard J. 1994. *Presidential Lightning Rods: The politics of blame avoidance*. Kensas: University of Kensas Press

Felstiner, William L. F., Richard L. Abel, and Austin Sarat. 1980. The Emergence and Transformation of Disputes: Naming, Blaming, Claiming, *Law & Society Review* 15(3-4): 631-654

Fischoff , B. 1975. Hindsight: Thinking Backwards, *Psychology Today* 8: 71-76.

Flatham, Richard E. 2002. *Thomas Hobbes: Skepticism, Individuality, and Chastened Politics*, Rowman & Littlefield

Gibson, James L., and Amanda Gouws. 1999. Truth and Reconciliation in South Africa: Attributions of Blame and the Struggle over Apartheid, *American Political Science Review* 93(3): 501

Hagiwara, Shigeru. 1992. The Concept of Responsibility and Determinants of Responsibility Judgment in the Japanese Context, *International Journal of Psychology* 27(2): 143-156

Hart, H.L.A. 1968. *Punishment and Responsibility*. London: Oxford University Press.

Haidt, Jonathan. 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment, *Psychological Review* 108(4): 814-34

Haidt, J.onathan, Bjorklund, F., & Murphy, S. 2000. Moral dumbfounding: When intuition finds no reason. Unpublished manuscript, University of Virginia

Heath, Chip, Richard Larrick, and George Wu. 1999. Goals as Reference Points, *Cognitive Psychology* 38: 79 –109

Hirschman, Albert. 1964. The Paternity of an Index, *The American Economic Review* 54(5): 761

Hood, Christopher. 2002. The Risk Game and the Blame Game. *Government and Opposition* 37 (1):15

Hood, Chrisopher. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton: Princeton University Press

----- 2004. Institutions, Blame Avoidance and Negativity Bias: Getting to Grips with the 'Blame Culture'. In paper presented at the CMPO Conference on 'Public Organisation and the New Public Management', March 19, University of Bristol

Hood, Christoper, Will Jennings, Ruth Dixon, Brian Hogwood, and Craig Beeston. forthcoming. Testing Times: Exploring Staged Responses and the Impact of Blame Management

Strategies in Two Exam Fiasco Cases. *European Journal of Political Research* 48: 695-722

Javeline, Debra. 2003. The Role of Blame in Collective Action: Evidence from Russia, *American Political Science Review* 97 (1): 107-121

Kahneman, Daniel, and Dale T. Miller. 1986. Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review* 93(2): 136-153

Kahneman, D., & Frederick, S. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.) *Heuristics and biases* (pp. 49–81) New York: Cambridge University Press

Kahneman, Daniel, and Cass R. Sunstein. 2007. Indignation: Psychology, Politics, Law. In *John M. Olin Law & Economics Working Papers*. Chicago

Kahneman, Daniel and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk, *Econometrica* 47: 263-91

Kohlberg. L. 1969. Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally

Kullback S. 1983. Kullback Information, in: Kotz, S. and N.L. Johnson (eds.). *Encyclopedia of Statistical Sciences* 4: 421-425. New York: John Wiley and Sons

Lewis-Beck, Michael S. 1988. *Economics and Elections: the major Western democracies*. Ann Arbor: University of Michigan Press

Leyden, Kevin M., and Stephen A. Borrelli. 1995. The effect of State Economic Conditions on Gubernational Elections: Does Unified Government Make a Difference? *Political Research Quarterly* 48(June): 275-300

Lin, Jianhua. 1991. Divergence Measures Based on the Shannon Entropy, *IEEE Transactions on Information Theory* 37(1): 145-151

Magurran, A.E. 1988. *Ecological Diversity and its Measurement*, Princeton: Princeton University Press

Malhotra, Neil, and Alexander G. Kuo. 2008. Attributing Blame: The Public's Response to Hurricane Katrina. *Journal of Politics* 70(1): 120-135

McGraw, Kathleen M. 1991. Managing Blame: An Experimental Test of the Effects of Political Accounts. *American Political Science Review* 85(4): 1133-1157

Molinar, Juan. 1991. Counting the Number of Parties: An Alternative Index, *The American Political Science Review* 85(4): 1383-1391

Mozaffar, Shaheen, James R. Scarritt and Glen Galaich. 2003. Electoral Institutions, Ethnopolitical Cleavages and Party Systems in Africa's Emerging Democracies, *American Political Science Review* 97(3): 379-390

Neto, Octavio A. and Gary W. Cox. 1997. Electoral Institutions, Cleavage Structures, and the Number of Parties, *American Journal of Political Science*, 41(1): 149-174

Oxford Advanced Learner Dictionary. 1989. Oxford: Oxford university Press

Powell, G. Bingham, and Guy Whitten. 1993. A Cross-National Analysis of the Political Context. *American Journal of Political Science* 37(2): 391-414

Rudolph, Thomas J. 2003a. Institutional Context and the Assignment of Political Responsibility, *Journal of Politics* 65(1): 190-215

Rudolph, Thomas J. 2003b. Who's Responsible for the Economy? The Formation an Consequences of Responsibility Attribution, *American Journal of Political Science* 47(4): 698-713

Page, Benjamin I. and Robert Y. Shapiro. 1983. Effects of Public Opinion on Policy, *The American Political Science Review* 77(1): 175-190

Schlenker, Barry R., Thomas W. Britt, John Pennington, Rodolfo Murphy, and Kevin Doherty. 1994. The Triangle Model of Responsibility. *Psychological Review* 101 (4): 632-652

Stimson, James A., Michael B. MacKuen and Robert S. Erikson. 1995. Dynamic Representation. *The American Political Science Review*, 89 (3): 543-565

Sulitzeanu-Kenan, Raanan 2010. Reflection in the Shadow of Blame: When do Politicians Appoint Commissions of Inquiry?, *British Journal of Political Science* 40(3): 613-634

Sulitzeanu-Kenan, Raanan and Chrisopher Hood. 2010. Blame and Blame Avoidance: Political and Psychological Perspectives. Unpublished manuscript.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton: Princeton University Press

Weaver, R. Kent. 1986. The Politics of Blame Avoidance. *Journal of Public Policy* 6(4): 371-98

Wilson, James Q. 1961. The Strategy of Protest: Problems of Negro Civic Action, *Journal of Conflict Resolution* 5 (September): 291–303

8. Appendix - Comparison of Coherence Measures

As explained in Chapter 5, measuring coherence is conceptually similar to the use of centrality measures of distributions, which are used as indicators in a range of scientific fields. Common examples include the Concentration Ratio (Cowling and Waterson, 1976: 271) the Herfindahl – Hirschmann Index (or HHI, see Hirschmann, 1964), used in economics to measure competitiveness; indices of party classification in political science, like the Taagepera and the Kesselman-Wildgen measures (Molinar, 1991); the Gini coefficient, Atkinson indices and other measures of inequality (Amiel and Cowell, 1999); a large number of measures of ecological diversity (for a comprehensive review see Magurran, 1988); different entropies and divergence distances (especially the Shannon Entropy and its derivatives, like the Kullback-Leiber divergence (Kullback, 1983, Lin, 1991)); and the kurtosis level of probability distributions. A distribution's variance can also be used to indicate how concentrated is the data relative to the mean.

To choose the appropriate measure to be used for evaluating coherence, we conduct a comparison of seven indices'[37] performance when applied to the measurement of agency coherence.

First, we compare the different values attained for a range of aggregate blame distributions with $|A^s| = 4$. The calculations were performed on the following distributions: [0.25,0.25,0.25,0.25],

[37] Variance, Herfindahl-Hirschman (HHI), Normalized HHI, Index of Dispersion, Kullback-Leiber divergrence (KL entropy), Gini coefficient, Kesselman-Wildgen Index (KW). Other surveyed indices were eliminated for being mathematically equivalent to one or more of these indices, and for relying on information irrelevant or missing from our model (especially the level of kurtosis, which, although computable, is inconsistent and lacks explanatory power because of not being applied to a single probability distribution).

[0.2,0.25,0.25,0.3], [0.2,0.2,0.3,0.3], [0.1,0.1,0.4,0.4], [0.1,0.1,0.1,0.7], [0,0,0,1], representing what we conceptualize to be increasing levels of coherence. They are presented in Figure 19:

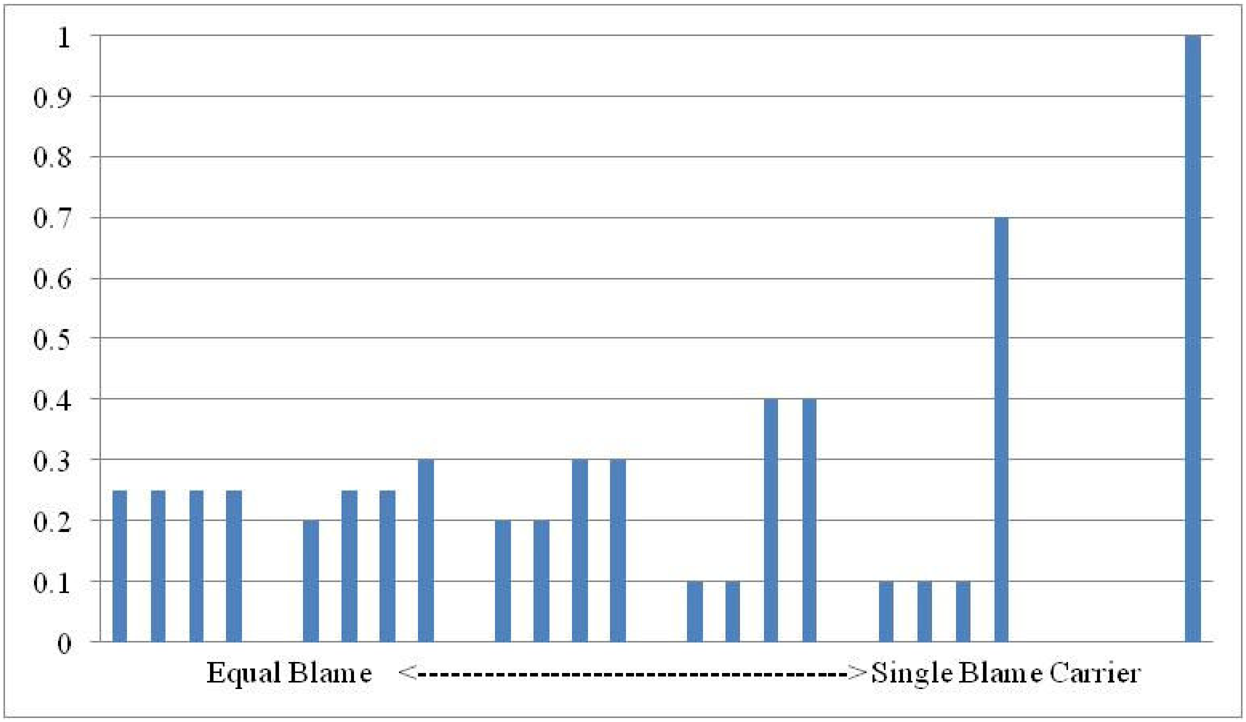

Figure 19 – graph representation, first set of tested blame distributions

The results for each measure are presented in Figure 20:

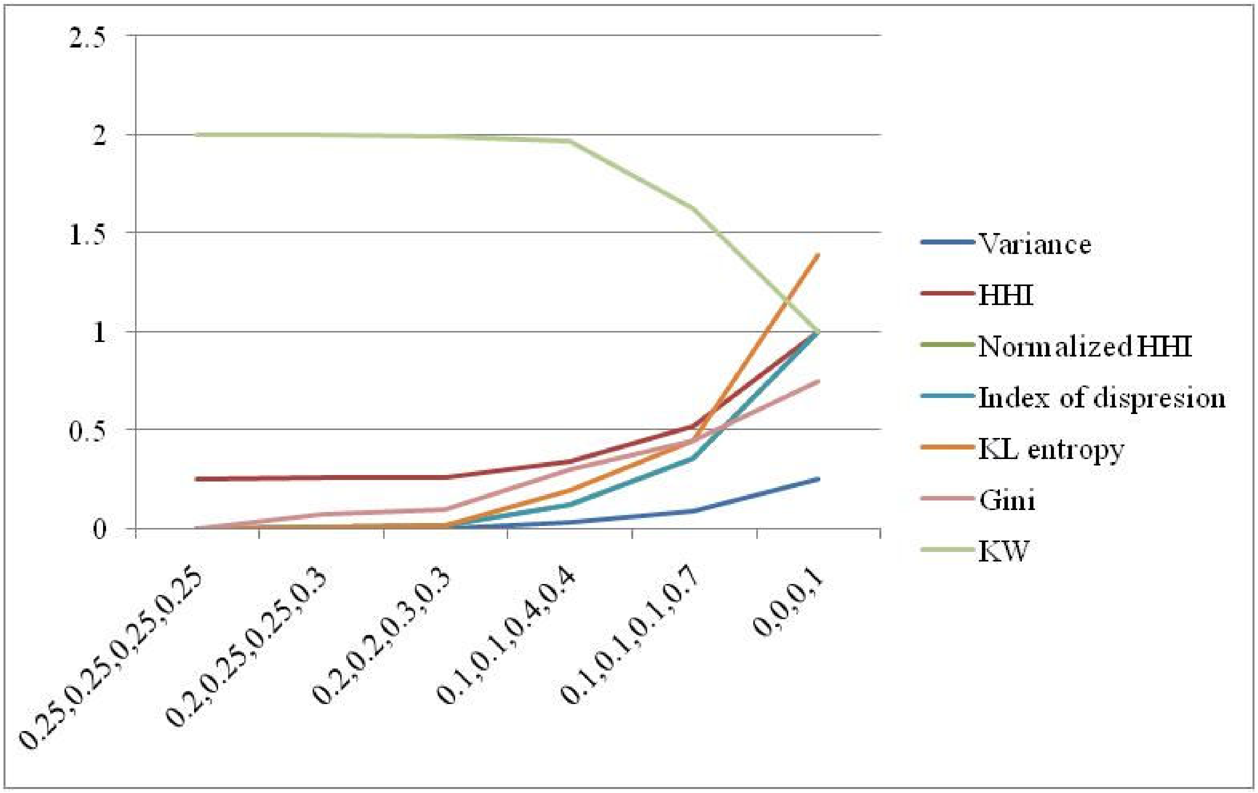

Figure 20 – measure performance for first set of distributions

Apart from the Kesselman-Wildgen index, all measures are highly correlated and return values consistent with our conceptualization of coherence. The KW index produces an inverse pattern which is virtually isomorphic to the Herfindahl values, and so is essentially equivalent as well. Most measures do not discern well between [0.2,0.2,0.3,0.3] and [0.1,0.1,0.4,0.4], apart from the KL entropy and the Gini coeffiecient, both displaying a stronger increase in value between the two distributions. The KL entropy also exhibits the most pronounced change in value when moving between [0.1,0.1,0.1,0.7] and [0,0,0,1], although we do not see them as profoundly different in coherence.

We now look at how these measures perform when applied to a second set of eight blame distributions with $|A^s| = 10$ for each set. They are displayed in Figure 21:[38]

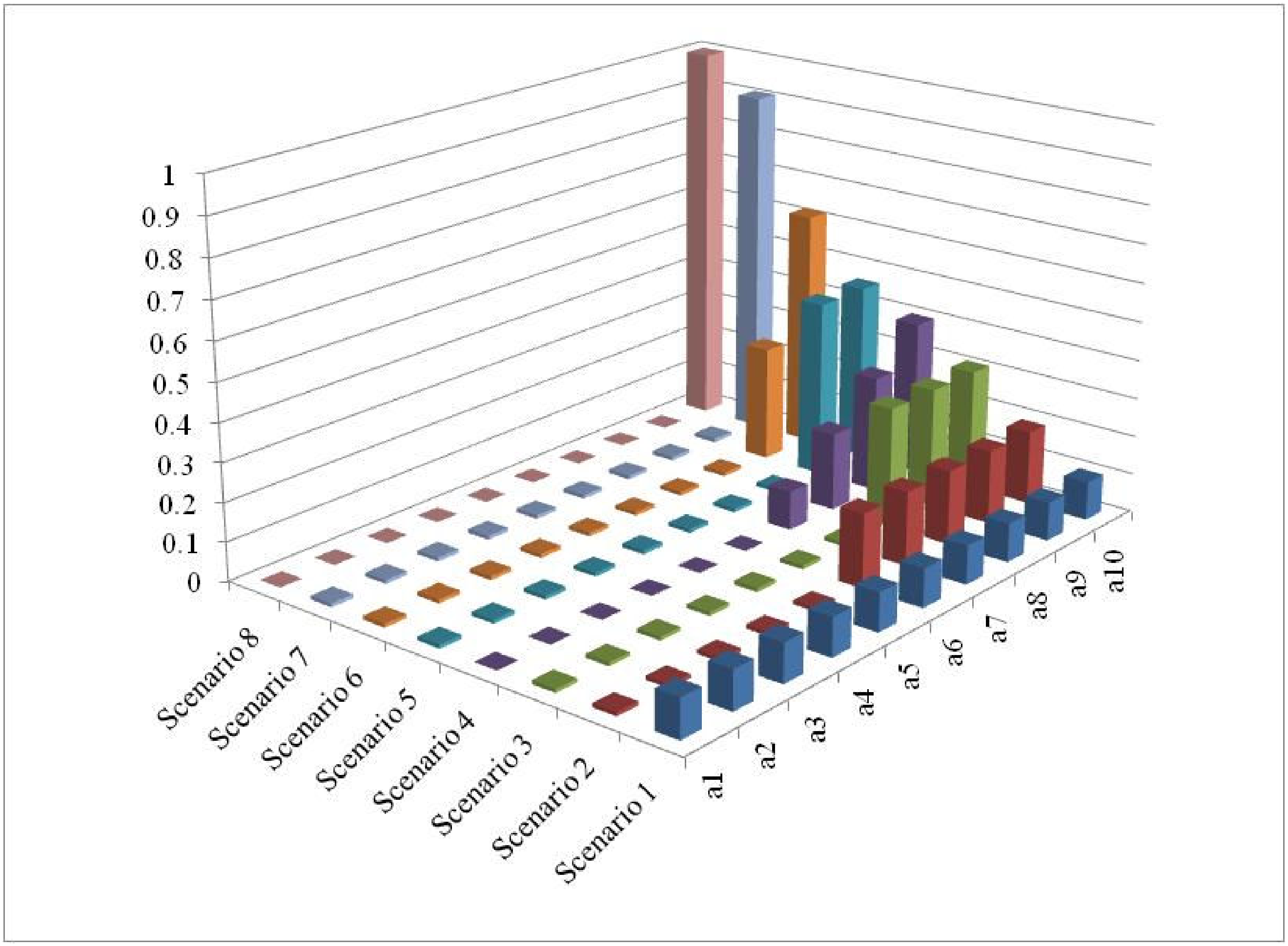

Figure 21 – graph representation, second set of tested blame distributions

[38] The exact aggregate blame values are:

[0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1], [0.01, 0.01, 0.01, 0.01, 0.01, 0.19, 0.19, 0.19, 0.19, 0.19],

[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.31, 0.31, 0.31], [0, 0, 0, 0, 0, 0, 0.1, 0.2, 0.3, 0.4],

[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.46, 0.46], [0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.3, 0.62],

[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.91], [0, 0, 0, 0, 0, 0, 0, 0, 0, 1]

The results for each measure are as follows:

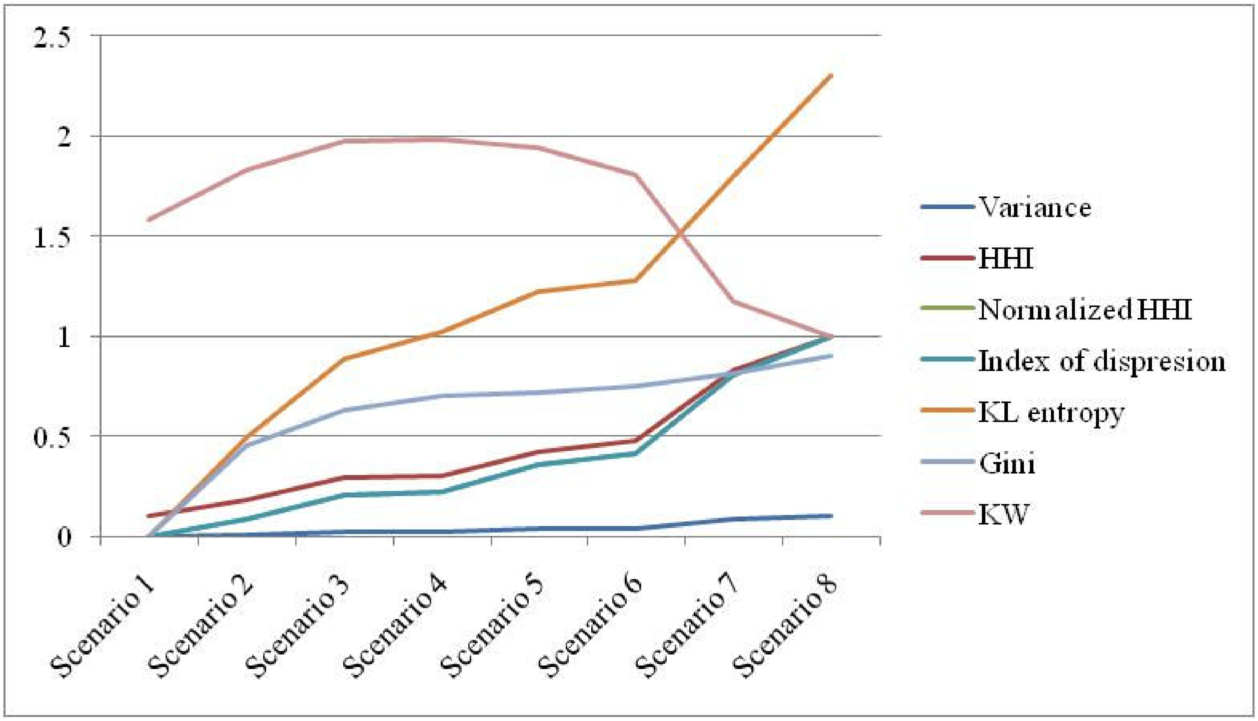

Figure 22 – measure performance for second set of distributions

There is greater variance of trends when increasing the number of agents. The Herfindahl indices and the KL entropy follow the same pattern (albit with different magnitudes), and variance as a measure of coherence yields exactly the same results as the normalized HHI (when we multiply the variance by 10 to attain the same magnitude). In contrast, the Gini coefficient fails to differentiate between scenarios 3 to 8, which involve decreasing numbers of blame-taking agents. This result follows from the small effect the dispersion of blame between the top 3-4 agents has on the Gini coefficient when all other agents receive zero or close to zero values. In this respect, the Gini coefficient is inadequate for the purpose of describing our notion of coherence, as we postulate that differences between blame-taking agents are important variables in the consolidation of aggregate coherence. Another index that is inconsistent with our conceptualization of coherence is the Kesselman-Wildgen measure: on the first three scenarios it provides results inconsistent with our expectations and with the other measures.

Finally, we compare how the four measures so far consistent with our conceptualization – HHI, Normalized HHI, variance and KL entropy - perform with respect to distributions of the type [0,…0,1] with a varying number of agents. We choose three aggregate blame vectors: [0,1], [0,0,0,1], [0,0,0,0,0,0,0,0,0,1], and expect coherence to increase as the number of agents rises, as this singles out more strongly the sole blame-taking agent[39]. The results are presented in Figure 23:

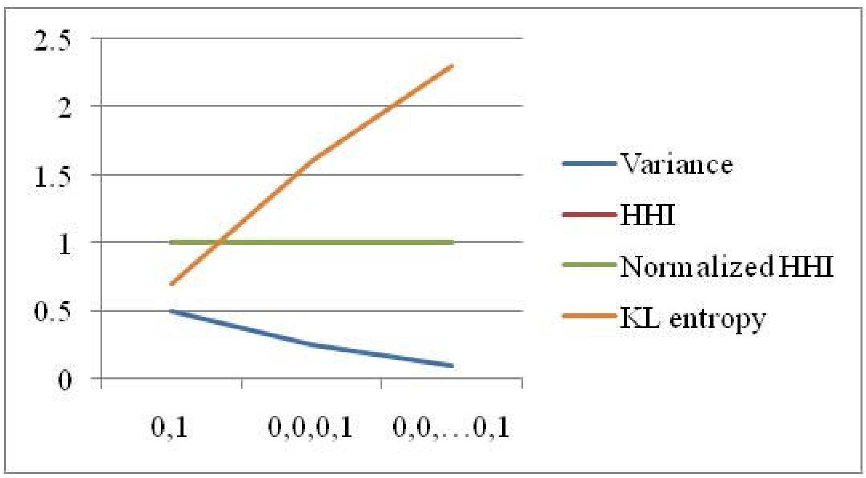

Figure 23 - measure performance for third set of distributions

HHI and Normalized HHI are indifferent to changes in the number of agents. The variance declines while KL entropy increases in value as more agents are involved in an issue, compared

[39] An agent's public blame is more pronounced if she is taking blame relative to a large group of non-blame taking agent, as opposed to being singled out as the blame carrier out of a small group of agents. This coincides with the intuition that it is less probable for an agent to be attributed with a large portion of social blame when there are many actors on an issue that could be perceived as contributing to the perceived loss.

with the same blame distribution over a smaller number of agents. Seeing that KL entropy is the only measure consistent with our conceptualization of coherence, we therefore choose a measure of aggregate agency coherence based on it.