constraints, and four note constraints related to the difficulty of tracing and attributing results to funds subsumed within a larger pool of funds not under evaluation. In the latter case, evaluations often assume that where the overall intervention achieved a particular impact, the contribution by the agency being evaluated supported that impact. Care needs to be taken to show this is a plausible assumption.

Discussion of constraints is not extensive. The potential constraint of carrying out affected population interviews with agency staff present is not commented on in most of the reports, although three of the DEC reports note the importance of interviewing beneficiaries without agency staff present (see Box 2.7).

Terms of Reference, Team Composition and Time Allowed

The studies provide almost no information about the process of developing Terms of Reference. It can therefore be presumed that these were issued by the commissioning agencies and generally accepted by the evaluation teams. This is in line with the 'conventional' approach adopted by almost all evaluation reports

Box 2.7 Good Practice: Participation, Beneficiaries and Gender

'As well as interviewing the agencies' project officers and key officials in coordinating agencies ... and partner agencies, a sample of beneficiaries will be selected and interviewed by the evaluators. These interviews will be conducted without agency personnel being present. ..The beneficiaries will be questioned on their views of the assistance provided the way they were selected and their overall views of the agency. Interviews with individuals may be complemented by discussions with groups of beneficiaries. So as to assess the agency's targeting and beneficiary selection methods the evaluation team will also interview a selection of potential beneficiaries who did not receive assistance.

It is expected that the evaluation team will use gender-aware and participatory approaches to seek the views of beneficiaries and, where appropriate non-beneficiaries.'

(DEC, 2000b)

(see Purpose and Focus of the Evaluation p53, Methodology and Transparency p58, and Chapter 3 p101).

Overall, Terms of Reference were adequate in directing the evaluation to cover key areas of programmatic focus (impact, connectedness, coherence, coverage, coordination, cost-effectiveness, efficiency, and relevance), as set out in the OECD-DAC (1999) Guidance. About three-quarters of the reports were organised around coverage of these areas. The OECD-DAC study, and papers and discussion leading to and from it, presumably had some influence in this respect, and there has also been an advance in methodological consistency since an earlier synthesis report (Borton & Macrae, 1997) which noted the lack of standards and commonality in Terms of Reference.

Terms of Reference were less adequate, however, in terms of directing evaluations to cover gender issues and to include representative samples from the affected population, both beneficiaries and non-beneficiaries. Presumably this is one of the reasons why limited attention has been paid to these topics. Furthermore, Terms of Reference did not generally specify the evaluation methodology to be used. While Terms of Reference may not need to specify the exact methodological approach to be taken, they should at least require adherence to good evaluation practice (for example, transparency in terms of the methodology used, triangulation, and adequate consultation with the affected population). A good practice case that provides general guidance but does not attempt to micro-manage, providing flexibility for the evaluation team, is given in Box 2.7.

Team composition also followed the conventional pattern, with most teams made up of consultants based in the country or continent of the commissioning agency. While there are potential advantages to this arrangement including the fact that evaluators may be a 'known quantity' to commissioning organisations, and that the pool of evaluation expertise may be larger in the commissioning countries, there are also disadvantages. These include the higher costs associated with the use of 'northern' consultants, the limited time they are able to spend on field work (often a result of fee rates and available resources); their frequent lack of appropriate language skills; and the loss of opportunity to build up indigenous capacity. Host country citizens, on the other hand, while not always having the breadth of experience or evaluation background of 'northern'

consultants, are likely to have language skills and sound local knowledge and contacts – invariably key factors contributing to an evaluation's success.

As with technical assistance, the choice of team members for evaluations of humanitarian action should complement and strengthen, rather than replace, indigenous capacity (Sobhan, 1997). Of the 33 individual evaluation reports, 21 were carried out by expatriates, 11 by a mix of host country citizens and expatriates, and only one carried out by host country citizens (DEC, 2000), the latter specifically to increase host country capacity. Mixed team reports – i.e., those by host country and expatriate staff – did not appear to be significantly different in quality to those carried out by expatriates only. This combination of team members may currently be the most appropriate team make-up, given that it provides both international and national expertise.

Only one-third of the reports note the expertise of team members. However, this generally only relates to sectoral expertise which makes it difficult to come to any overall conclusions about whether evaluation teams included relevant personnel and skills, including language skills (see section on Legibility below).

Another feature in nine of the reports (three by NGOs, five by UN agencies, and one by a bilateral donor) was the inclusion of agency staff on the evaluation team, though none discussed the implications of this. While this may increase participation and the likelihood of uptake of recommendations, it may also be seen as a method of biasing results and so lower the credibility of the study. Reports ought to explicitly state the advantages and disadvantages of including agency staff (particularly where the staff come from outside the agency evaluation office) as 'internal' evaluative processes are often viewed as less impartial than those carried out externally. As is well known, negative evaluation results are subject to extensive discussion and negotiation, with evaluation offices often playing the role of mediator between the evaluation team and the department or agency under review. This process itself can lower the credibility of the evaluation because it may appear to call into question the evaluators' impartiality. It is therefore particularly important to provide a clear rationale when including agency staff – particularly as inclusion of agency staff is not a practice that is common in the general evaluation field[7].

Almost none of the reports note how their evaluation team was selected, whether through tender, head-hunting or, in the case of mixed teams, how internal members of staff were selected. Evaluation reports could add to their credibility if this process and any constraints involved in team selection and composition were noted.

In assessing whether adequate time was allowed for the evaluation, it was presumed that teams would have sufficient time for developing the methodology, reviewing background documentation, consulting with the affected population and agency and government staff, writing and disseminating a draft report, and finalising it. Time for these functions is considered sufficient in 13 of the studies, and insufficient in 15; five studies do not note the evaluation's timing.

Overall, consultation with the affected population did not produce adequate information to support impact assessment, or the analysis, in many evaluation reports. For the most part this appears to result from insufficient time being allowed for consultation, and there would seem to be a direct correlation between the quality of an evaluation and the allocation of adequate time to consult with the affected population.

With regard to the allocation of time, there do not appear to be any systematic standards directing commissioning agencies in this area. For example, a team evaluating US$16m in expenditure was allowed about 200 person days (DEC, 2000); a team evaluating over US$100m and the establishment of a new state appear to have spent about 120 person days (UNTAET, 2000); and a team evaluating US$606m appears to have been allowed about 60 person days (WFP, 2000a)[8]. Time allocated may be determined by both commissioning agency requirements and the time that evaluators have available. When expatriate team members cannot spend adequate time in the field, host country team members may have more time available and be able to fit fieldwork around other commitments more easily.

Information on Context and Intervention

Provision of contextual information in relation to the humanitarian intervention proved to be one of the stronger areas of the reports – 29 provide it. Analysis of economics and agriculture is stronger than geopolitics and social issues, although the latter are mostly adequately covered. Cultural issues, however, are not discussed in any depth, which means that there may have

been little understanding of the ways in which culture (for example, intra-household relations, and 'community' or ethnic relations) has affected and shaped humanitarian action.

Only six reports attempt to construct a narrative history. Of these, four include all relevant actors (agency and government staff and the affected population), one includes agency staff and the other includes local level officials. This also reflects the conventional evaluation approach of most of the studies. However, as DAC-OECD notes: 'In attempting to understand situations and structures, to analyse a particular set of events and processes the construction of a narrative history is a powerful tool' (DAC-OECD, 1999: p18). Specific skills – in particular the ability to listen and sympathise while remaining impartial – and adequate time are necessary for this challenging exercise, in particular, in discussion with the affected population. This is thus another area where commissioning agencies need to provide a clearer lead

Eighteen reports provide adequate reference to secondary reports, and 12 to comparable reports, such as similar studies from another country, region or sector. Finding a balance here was difficult for some evaluators, whose reports read more like PhD theses than evaluation reports, with, for example, extensive historical detail or description of the intervention. Future guidance on evaluation methods should provide details on what can be considered adequate in this area.

Methodology and Transparency

The 33 individual evaluation reports were also assessed in terms of the extent to which their methodology met what is generally accepted to be good practice. Rather than the 'methodological anarchy' mentioned in Borton & Macrae (1997), the reports showed considerable consistency in terms of their following what has been termed the 'conventional' approach This approach – that is, the use of different information sources (in this case interviews with agency/government staff, the affected population in some cases, and review of documentation) – might be considered as the mixed-method technique currently promoted in general evaluation theory as a strong evaluation approach. However, few reports cross-checked or made comparisons between these different sources of information. Very few experimented with other techniques, such as self-assessment. Even techniques now commonplace in the development field, such as Participatory

Rural Appraisal, were rarely used (exceptions included DEC, 2000 and WFP, 2000; see Box 2.8) The evaluation of the intervention in East Timor (UNTAET, 2000, 2000a, 2000b) did involve a mix of techniques: a self-assessment, a study of beneficiary participation, and an external evaluation. It thus had an opportunity to triangulate, but this does not appear to have been done.

Almost all reports are based on the concept of the external 'objective' evaluation team, and thus have not taken on board some of the advances in the evaluation field in the last decade. These include a focus on greater participation of all actors in the evaluation process, greater attention to highlighting good practice in operations, and effective use of evaluation results. In addition, evaluations and their methodological approaches, including those that attempt to remove 'subjective' elements, are always influenced by the perceptions and biases of the evaluators. This is clearly reflected in the reports reviewed in terms of their gender blindness and, for example, that some evaluators thought it more important to cover geopolitics or consult with beneficiaries than others. To increase rigour and transparency, evaluators need to make clear how their perspective has influenced their methodological approach (see Chapter 3 p101 'Opaqueness of Evaluation Methodology', for a typology that can guide this exercise).

As noted, most reports cover the recommended evaluation areas

Box 2.8 **Good Practice in Use of Participatory Techniques and Control Groups**

'Methods included stakeholder workshops, village PRAs, gender analysis, focus group meetings, semi-structured household interviews, and spot visits to FFW sites. Participatory tools included key informant interviews, use of ZOPP-type cards, participatory mapping, calendar making, matrix ranking of FFW activities, and positive/negative impact analysis. Gender issues were mainstreamed throughout the exercise ... To compare the "with" and "without" project situation as well as beneficiaries and non-beneficiaries, mini-PRAs were undertaken in eight severely affected villages, five moderately affected villages and four non-project villages.'

(WFP, 2000)

of impact, appropriateness, and so on, with the notable exception of cost-effectiveness. This is covered in only nine of the reports, and while it may be generally felt difficult to capture (OECD-DAC, 1999), where possible an assessment of cost-effectiveness will provide useful information on whether expenditure was adequate While these 'standard' areas are usually covered in adequate depth, the quality of the information sources used raises some concern – for the most part these sources are agency staff and agency documents. Commissioning agencies need to make it clearer, when providing Terms of Reference and guidance to evaluators, that as wide a range of information sources as possible should be used, with triangulation between different sources being an essential methodological element. Not only will use of a wide range of information sources strengthen evaluation methodology, but it will also provide greater credibility, as it adheres to evaluation good practice.

Only one evaluation used a control group approach[9] – a striking finding given that quasi-experimental design was required by some of the Terms of Reference and is increasingly standard practice in development cooperation evaluations. This is one of the weakest methodological areas in the studies and requires attention from commissioning agencies.

Consultation with Beneficiaries and the Affected Population

OECD-DAC notes that: 'Experience shows that interviews with beneficiaries can be one of the richest sources of information in evaluations of humanitarian assistance' (OECD-DAC, 1999: p25). Despite the considerable emphasis placed on, and experience gained in, this area over the last decade, this set of reports is weak in this area. Only four reports consulted an adequate number of beneficiaries and clearly note the methodology used for this consultation. 13 studies did not consult with the affected population. There is only circumstantial evidence of consultation in the remaining 15[10]. For example, one evaluation report notes that a decision was taken to consult with beneficiaries, but no further information is provided about this; in others beneficiary comments are dotted throughout the study to support particular statements, but no general methodological details are given.

However, it must be said that evaluation teams found consultation with the affected population easier or more relevant in some

situations than others. For example, evaluations that cover several years, such as the DANIDA reports, had difficulty in locating beneficiaries from the start of the period being evaluated. In other instances security of evaluators was a major issue. In most cases, however, it was possible to consult with a sample from the affected population and there does not appear to be any reason why evaluation teams who have done this should not clearly note the methodology they used. In those cases where consultation is already institutionalised (for example, the DEC), commissioning agencies could go further in ensuring that the affected population also plays a role in the evaluation process, and that results are circulated to consulted groups.

Attention to Gender and the Vulnerable or Marginalised

Only one-third of the reports contained findings on gender, or could be considered to be partly gender mainstreamed. However, even in these reports we learn very little about either the effects of complex emergencies and disasters, or the impact of humanitarian action, on gender relations and gender equality. Reports that do cover gender tended to confuse gender with women, with the result that gender equality was not considered. For example WFP (2000a) quotes a nutrition study which showed that boys in the 6–24 month age group were nearly twice as likely to be physically wasted as girls; the section on 'Gender Issues', however, refers exclusively to women rather than to relations between women and men. The 22 reports that do not cover gender adequately either do not mention gender issues or cover them briefly in a separate section of a few lines.

The reports that do cover gender adequately illustrate that it is an important area and one that it is possible to include in evaluation consultancies. These reports note that women in the affected population play an active role in the relief and rehabilitation phases (a point also made in WFP 2000), and that gender insensitive interventions during these phases can damage longer term efforts aimed at promoting gender equality. In most reports, however, gender is absent, and even basic norms, such as disaggregation of statistics by sex, are not followed. Given that gender, as noted in OECD-DAC (1999), is central to our understanding of complex emergencies and the effectiveness of humanitarian action, it can be deduced that

evaluators and commissioning agencies are failing in this area. There may currently be increasing attention to gender issues, for example in the 2001 UN consolidated appeal process (UN, 2001) This appeal process was an attempt by OCHA to integrate gender questions into the CAP, for example, by focusing on women and war.[11]

The reports reviewed fare better in their attention to vulnerable groups in general, with 20 paying adequate attention to groups such as children and the elderly. Once again, however, there is a tendency not to disaggregate and to discuss the 'vulnerable' in general terms. Greater disaggregation would allow a clearer understanding of the intervention's impact.

Coverage of Factors Potentially Influencing Performance

The two areas considered here concern contextual and organisational factors. In terms of the former, about two-fifths of the individual reports can be said to cover adequately the geo-political context, relations with local authorities and/or the national government, and diversion of assistance away from humanitarian programmes and its potential use by militias and militaries. Given the importance of diversion and corruption in public debates on humanitarian action, commissioning agencies may want to ensure that more systematic attention is paid to this area. Very few studies consider social and cultural factors that might have been a hindrance to intervention impact (for example, intra-household relations or food preferences).

More of the reports cover organisational factors. In particular, assessment of management of the intervention cycle, from design through evaluation, is a strong point of the studies. Implementation and monitoring, for example, are adequately assessed in almost all of the reports. There is, however, somewhat less focus on design and evaluation. Head office/field relations are covered in about one-third of the reports.

Conclusions and Recommendations

This part of the proforma allows assessment of whether key findings have been shared with stakeholders, the logic of flow between a reports' findings and conclusions, and between the conclusions and recommendations, and the apparent 'implementability' of the recommendations.

Twenty-one of the reports reviewed had shared draft versions with stakeholders, usually agency personnel and donors. In the other 12 cases there were either no details on this topic or the draft reports had not been shared The fact that a majority was shared is encouraging and suggests that consultation on the draft is beginning to become institutionalised, at least at the agency level. However, sharing of findings with the affected population was almost non-existent.

The development of clear conclusions and recommendations, that are linked to the main text and to each other, is a strong area in the reports with just a few exceptions. Evaluators were skilled in summarising their findings and drawing recommendations from these The level of expression of recommendations varies: 14 studies include recommendations that are specific and include details in relation to implementation, 10 include quite general recommendations, while nine reports either include no recommendations or an unclear set of recommendations. As DAC-OECD (1999) notes, there are no hard and fast rules in the writing of recommendations. Given the diversity of evaluation reports assessed it was not possible to come to a conclusion as to whether evaluators were managing to fit their recommendations to the needs of the commissioning agency or not.

Legibility and Dissemination of the Final Report

The proforma also allows assessment of the clarity and accessibility of the final report in recognition of the potentially diverse readership, and whether the principal lessons are identified in an accessible fashion.

Overall, 25 of the reports are assessed as clearly written and accessible, and 15 as making good use of maps and diagrams. Twenty-eight reports clearly identify the evaluation's principal lessons in an accessible fashion, and the reports are generally strong in directing different messages to different audiences, through use of the Executive Summary and lessons learned. In some reports the writing style was unclear, and greater attention to editing would have made for a much stronger evaluation, for example, in the wording of recommendations.

CHAPTER 3

EVALUATIONS OF HUMANITARIAN AID IN RESPONSE TO THE 1999 KOSOVO CRISIS: SYNTHESIS AND META-EVALUATION

CHAPTER 3

EVALUATIONS OF HUMANITARIAN AID IN RESPONSE TO THE 1999 KOSOVO CRISIS: SYNTHESIS AND META-EVALUATION

3.1 Overview of this Chapter

3.1.1 Introduction

In June 2000, aware of the opportunity provided by the sizeable emerging cluster of evaluative reports of the humanitarian response to the 1999 Kosovo Crisis, ALNAP commissioned an analytical overview (synthesis and meta-evaluation) of evaluative reports of humanitarian action undertaken within this single context. That original analysis, prepared by Professor Raymond Apthorpe, provided the background paper[1] for the ALNAP October 2000 Symposium[2]. This chapter expands on that analysis to reflect issues raised by symposiasts (including ALNAP members) and incorporate evaluative reports published since.

Its twin tasks are to:

1. summarise the principal findings and recommendations by means of synthesis;

2 assess the quality of the reports through meta-evaluation.

Section 3.2 'The Kosovo Conflict: a Set of Interlocking Crises' identifies and reviews problems presented by evaluators as having reached crisis proportions, and therefore their principal foci – whether explicitly stated in Terms of Reference or introduced by the evaluators themselves. The premise being that the first task of an evaluation or overview of evaluations is to discover how the problem was perceived and review in consequence the appropriateness of indicators used to validate the response.

Section 3.3 'Agency Response Issues' selects and synthesises recurring themes in the evaluations' findings, conclusions and recommendations under the broad categories (a) humanitarian principles, practices, and strategic planning, (b) human resource organisation and management, and (c) technical and standards aspects of operations. It seeks to evaluate the strengths and

weaknesses of programme design and implementation, based on the premise that evaluation should be multifaceted and consider the wider socioeconomic and political context. The tendency when considering appropriateness of response, for example, is for humanitarian evaluations to focus only on donor relief-supply factors, leaving issues such as agency-driven supply and ideological agendas outside the critical frame.

Section 3.4 'Meta-evaluation: Kosovo Evaluations as Learning Tools' evaluates the Kosovo evaluations to assess the generic practice of evaluation of humanitarian action, and its strengths and weaknesses as a lesson learning genre.

Throughout this chapter the term 'programme' is used inclusively, denoting both policy and operational aspects within humanitarian action, as is the phrase 'organisational/institutional learning' denoting both the skills and processes associated with knowledge management and implementation.

3.1.2 Characteristics of Evaluative Reports Considered

The international response to the 1999 Kosovo conflict ranks as one of the largest, in terms of the scale of resources involved, which in turn has generated a multitude of evaluative reports.

The international response to humanitarian needs in Kosovo has been the subject of at least 20 separate evaluations of humanitarian action of which 16 plus 1 synthesis have been published or made available to ALNAP.

While the central findings and recommendations of this chapter are primarily in respect of the ALNAP Evaluative Reports Database (ERD) 'evaluations' category, to increase the richness of analysis, it also draws on 50 or so other types of evaluative reports, including 'After Action Reviews' conducted by NATO and its member states, parliamentary committees of enquiry, scholarly reviews and seminar reports.

It is the perception of this chapter's principal author that most categories lack discreteness and most types of report lack serious social analysis. Their management-solutions approach exhibits stakeholder bias that excludes beneficiary perspectives. The sample reports, with the exception of a few French ones, have been made available in English with no evidence of the availability or even existence of other foreign language evaluative reports of the humanitarian response to Kosovo.

A characteristic, particularly in evidence in the 'evaluations' category, is the extent to which like evaluated like, with humanitarian agencies engaging 'ex' or 'current' humanitarian practitioners, and the military, other military.

Most evaluators, although experienced in the undertaking of humanitarian evaluation, had no prior experience of the Balkans.

An unusually large number of the reports focus on single (or umbrella) agencies or cross-cutting themes, but there is no system-wide evaluation and only one joint donor/agency (UNICEF/DFID) evaluation, of limited scope. The lack of a system-wide evaluation, although not uncommon, is unfortunate since what we have as a sample is a collection of uncoordinated and, to an extent, overlapping reports.

As one evaluation notes: '[T]hough rarely conducted, joint donor/partner agency evaluations are feasible .. [A] multi-disciplinary team including both "insiders and outsiders" ... can produce a healthy mixture of perspective, knowledge and experience' (UNICEF/DFID, 2000: p41) Certainly a series of specialist sectoral analyses, thematic explorations, and so forth, broken down by emergency phase, locale and principal actors, would be invaluable.

The International Crisis Group's (ICG) *Kosovo Report Card* (ICG, 2000), a review of the international mission's first fourteen months in Kosovo, although not an evaluation *per se*, comes closest to providing a system-wide overview, and yet, has a limited focus on humanitarian issues, is highly selective and, finally, rather abstract despite some illustrative detail in two of its fifty pages.

However, individual agencies are probably still of the view that they can learn more from an evaluation focused on their actions, and the current reality is that there is no appropriate system-wide institution for such evaluations.

Lastly, despite the majority of the reports being on the ERD and within the public domain, several remain restricted. Such practice does not sit well with the ethics of transparency and accountability of non-profit organisations and international agencies.

3.1.3 Approach and Methods: Synthesis & Meta-evaluation

As Roland Barthes put it, 'the meta-book is the book that talks about the book'. Yet 'effective synthesis ... [is] not afraid to go

beyond the data as necessary' (Perrin, Presentation to the UK Evaluation Society's Annual Conference, December 1999). This is the domain of this chapter: 'proactive[;] ... not afraid to improve, not prove' (ibid.).

Every approach to critical analysis has a direct influence on its outcome. As is common in the logic of evaluation theory and practice[3], the approach adopted here is one that aims to be as inductive as possible, working up from the material being considered rather than down from a deductive, theory based, schema. Although not mutually exclusive – and particular cases, discourses, paradigms, and comparisons get drawn on by both – the advantage of the inductive is that of not having to force material into pre-conceived grooves, nor having to pass it over if it doesn't fit. It also allows the adoption of different modes. This paper is somewhat unusual in that 'aid' studies tend to be heavily modernist rather than post-modernist, and doctrinal rather than explorative of alternative thinking and positions, in their social science approach.

3 1 4 Credibility of the Data Sources Considered

The value of the synthesis of the conclusions, findings and recommendations presented in this chapter is heavily dependent on the quality of the evaluation reports themselves.

Unlike most of the other evaluative reports considered, the evaluations were undertaken by evaluators experienced in evaluating humanitarian action, and feedback would indicate that the reports were generally well received by their commissioning agencies.

However, despite the fact that the credibility of 'evaluation of humanitarian action' as a genre is no longer challenged, the effectiveness of current practice, for which commissioning agencies and evaluators share responsibility, is to be questioned.

The most common approach to evaluation is one of short-term consultancy, often paid for by the agency being evaluated with the objective of making recommendations on how to improve subsequent performance. Unlike social research, management consultancies tend to proceed directly to bottom-line judgements on the issues, often narrowly defined by commissioning agencies in their Terms of Reference.

Recommendations emerging from such consultancies, although generally well spelt-out, come without qualification or allowance

for margin of error. They make no acknowledgement that 'fundamentals forgotten' may be 'mistakes made' or vice versa, and provide no scenario of options with accompanying positives and negatives. Instead, a classical 'single best' business solution is laid out in the 'big idea' tradition of management consultancy.

The majority of evaluations pay scant regard to serious basic description or categorisation, providing broad-brush critiques without in-depth analysis. Where do the militia fit in the criticisms of military breaches of humanitarian principles? Basic chronologies are uninformed by political economy sensibility or social analysis, and one evaluation report even notes that there had been insufficient time allowed for headquarter enquiry.

The comparative advantage of the evaluation genre may not be for organisational learning (in-house workshops may do this better), and its present standing as a tool for social learning leaves much to be desired. But, what can or should one expect from a relatively short-term consultancy by someone with no previous experience or knowledge of the area. More importantly, are evaluations in their current form sufficiently informed to be reliable?

Nevertheless, for all the current problematic in respect of practice, quality and resourcing of evaluations of humanitarian action (perceived and real), the story of the humanitarian response to the Kosovo Crisis that emerges here draws particularly from the evaluation genre.

The extent to which lessons learned from previous experience fed through into action in Kosovo remains to be seen. Despite evidence that lesson learning *does* happen: '[T]he lessons learned in ... Bosnia has [sic] permitted ECHO to avoid a catastrophic outcome [in Kosovo]' (ECHO, 2000o: p9). This is heavily countered by the emphasis in every evaluation that the old lessons have still not been properly learned.

3.2 The Kosovo Conflict: a Set of Interlocking Crises

3.2.1 Introduction

Every effort must be made to ensure the emerging story of a complex humanitarian response to a complex emergency is not oversimplified in the telling. The story drawn from these evaluations is multi-faceted, one evaluator even questions whether agencies were responding to the same conflict. There were numerous interpretations and perceived crises, where programmes placed different emphases on the same aspects.

Just as the various parties to a complex emergency understand it through the lens of their own interests and organisation, so it is helpful for this paper to suggest an analytic framework from which to interpret this humanitarian response. Adopting a post-modernist approach allows 'the same' emergency to have 'different' meanings requiring 'different' programmes of action to meet differences in vision, standards, and objectives. This section considers the various crises as perceived by the different actors.

Some reports find it to be a sudden onset type emergency, the precise or even approximate dimensions of which could not reasonably have been anticipated due to scarce, non-existent or illegible early warning signs. The majority, however, takes the diametrically opposite view that its onset was far from sudden and that it could, and was, clearly to be seen by anyone looking seriously.

Other reports consider whether this was a conflict that NATO *had to have*, but important as this debate is, the focus of this section is on issues that can be addressed outside it. Given the focus on humanitarian, not military, 'success', this section's concerns extend only to NATO's relatively important role in relief work and whether it signals an end to humanitarianism as we know it.

3.2.2 A Crisis of Humanitarian Principles

What distinguishes this particular humanitarian action from many others is the extent to which it is dominated by the dilemmas and paradoxes thrown up by NATO's involvement – particularly since those governments sending in bombers were also funding humanitarian efforts. Programmes and evaluations alike recognise that NATO's involvement in the overall sequence of events was huge and decisive.

The evaluation reports broadly see NATO's involvement as an actual or potential contradiction of humanitarian principles, and

its actions as presaging a new humanitarian order in which cowardly political and diplomatic action on the part of external actors has led to 'humanitarianism by default'. But the evaluations also recognise the contribution made by NATO, directly or indirectly, to the effectiveness of the humanitarian operation (including rapid refugee camp construction, despite some poor siting decisions). One or two evaluations note the military's need for good civil–military/military–civil relations, as well as issues of relative competence and efficiency. However, the analysis of civil-military relations is limited primarily to NATO, excluding for example INGO–militia relations

It is fascinating therefore to learn that when the conflict was at its height, discussions concerning humanitarian principles such as neutrality, impartiality and proportionality could, with notable exceptions, be heard more in military than civil circles. As was, and seemingly still is, the case in respect of defining procedures for civil–military relations.

3.2.3 A Multi-ethnic Society in Crisis

Like the programmes they evaluate, the evaluation reports perceive the Kosovo conflict as 'a multi-ethnic society' in crisis. Only one complains about the distortion and caricaturing by western commentators that reduced the complexities of the 1999 Kosovo crisis to one of a tinderbox of Serbo-Albanian tensions ignited by and for the benefit of Milosevic's ambition. Virtually none mention, let alone consider, the different meanings of ethnicity (or, for that matter, multi-ethnicity) as the affected populations themselves understand and act on them.

Like the media commentary, the evaluative reports fixate on ethnicity as for the most part essentialist – primordial *identity*. There are no perspectives on ethnicity understood or interpreted as, say, *image* (that may have been fabricated), or *badge* (that can be pinned on or taken off), or anything else. The reports approach the subject as if all Balkan politics are identity politics, and ethnicity not just one cultural and social marker among others – or, for example, an indicator less of specifically social or cultural distinctiveness than a symbol of geopolitics or a token of history and place

This weakness with regard to ethnicity and other basic sociocultural, political and economic institutions has serious consequences. Little appears to have been learned from, for instance, Johan Pottier's work[4] on ethnicity undertaken following

the Rwanda crisis. It makes it difficult for an induction-led synthesis and meta-evaluation to critique the evaluations and other evaluative-type reports effectively on this point, and has major implications for team composition as well as Terms of Reference writing.

3 2.4 A Crisis of Protection

A critical aspect of this multiple crisis perception is whether programmes and their evaluations saw the emergency primarily as one of assistance or one of protection. 'Broadly, DEC agencies have been less animated about protection issues in Kosovo than they were during the refugee crisis ... A lack of INGO activity in human rights and protection has (ironically) drawn criticism from one donor, which argues that NGOs' responses have been largely relief driven and financially driven' (DEC Vol. II, 2000c: p108).

Many of the policy review studies and general media reports tend to see crimes against human rights as constituting the fundamental crisis in Kosovo, whereas most of the evaluations, as the programmes they evaluated, focus not on protection but on assistance. This may be partly because at the time of the NATO bombings, when most needed, the humanitarian agencies simply were not there, but it may also reflect perceptions of mandate. For instance: 'One reason why NGOs may have been inactive on protection issues in Kosovo is the presence of other organisations who are better able to provide physical protection (KFOR and UNMIK police) or more experienced in monitoring and reporting, and with a special protection mandate (UNHCR, OSCE and ICRC)' (DEC Vol. II, 2000c: p108).

For the most part evaluation Terms of Reference did not specify protection as a point of focus, an omission seemingly unquestioned by evaluators. Should not this uncritical acceptance of Terms of Reference, and the failure to look at agencies' broader objectives and policies to identify any protection mandate therein, be a point of concern?

Despite the absence of specific guidance for the evaluation of protection, these evaluations omit it at their peril. If saving life is part of any definition of humanitarian action, it is simply not credible to ignore the issue of protection (including armed-protection) in an analytical account of humanitarianism. Whether or not military involvement compromises humanitarian principles or service, the danger of excluding armed-protection from the framework and analysis of the humanitarian system is

that it will be seen as, and/or become, the sole responsibility of actors with non-humanitarian objectives, priorities and agendas.

It is fair to say that complementary strategies, such as advocacy and affirmative action with protection objectives, were successfully used by some agencies to address the realities of 'organised criminal violence, gender based violence, general property rights and the property rights of women in particular.' (DEC Vol. II, 2000c: p107). But only the type of humanitarian protection achievable without the use or threat of armed force was pursued by agencies: '[D]uring the crisis most [of the DEC-supported agencies] did develop and employ protection measures at different times and in different places, but for some this was not a conscious strategy' (DEC Vol. II, 2000c: p101).

Curiously, it is again primarily the military (NATO) reports that address the protection of vulnerable minorities, while the non-military reports tend to make general statements about physical security of persons and groups under threat. This characteristic is particularly noticeable and limiting in any discourse about conflicts of principle and mandate between military and civil actors.

In addition, in a context where human rights violations constitute the crisis, the use of famine and morbidity indicators for humanitarian programmes is highly questionable. The comment that '[N]ot a single Kosovar seems to have died from lack of food, shelter or emergency health supports, which provides a very clear indicator of actual achievements' (ECHO, 2000q: p5), clearly fails to address the obvious question of how many deaths occurred due to lack of protection.

Overall, the story that emerges is one of (relative) survival *in spite of* the humanitarian effort. One condemnation reads as follows: 'Agencies' decisions to withdraw from Kosovo were based on their assessments of security and their ability to continue working. Some point out that NATO prevented humanitarian action from the air, while the Serbs prevented it from the ground. However by withdrawing en masse humanitarian agencies, including DEC agencies, effectively failed to sustain "*humanitarian space*". In the words of ICRC the "*black hole*" in Kosovo was "*a collective failure of protection*" ' (DEC Vol. II, 2000c: p106).

3.2.5 A Relief-supply Crisis of Spend, Spend, Spend

It is not unusual for problems to be attributed to donor-driven, relief-supply factors in humanitarian evaluations, and this set of

evaluations is no exception While several note that the overall humanitarian response in Kosovo was driven by geopolitics (predominantly those of the US and UK) as well as NATO's own agenda, a fundamental problem was that of 'too much', not 'too little', aid: '[O]ne distinct characteristic about the aid effort in Albania over the first three months was the apparent absence of ... any awareness that resources were finite. There was money to do almost anything and to do it almost anywhere' (Porter, 1999: p22).

A related, and frequently noted, aspect of the response was its strong bilateral nature with donors bypassing multilaterals to contract responsibilities to their 'own' national INGOs. This resulted in 'uneven standards set by the bilateral actors ... [which] in some camps . . [were] so high that people, only half jokingly, came to ask whether there should be maximum standards as well as minimum ones' (Porter, 1999: p22). The majority of INGOs appear to have accepted this lack of impartiality and neutrality with varying degrees of embarrassment, whilst apparently doing little or nothing to reduce or change it.

3.2.6 A Crisis of International Unpreparedness

A lay-person, knowing little of the humanitarian system, might be forgiven for assuming that any major regional/international emergency, particularly where the big powers had a stake, would benefit from a standing international rapid emergency response capacity. What emerges from the Kosovo reports is that no such capacity currently exists.

The question therefore arises as to whether our current international humanitarian relief system seriously qualifies as such if it lacks instituted readiness for immediate response, including a capacity for emergency assessment reconnaissance. In the past, this lack has been blamed on cost factors, but material circumstances alone are unlikely to have carried the day. Questions relating to mandates, organisations, rules, regulations and discursive practices must constitute other factors.

3.2 7 A Crisis of Leadership and Coordination

The multi-faceted crisis of the multi-faceted concept of 'coordination' is a crisis of the humanitarian system's own making. The Kosovo evaluations are particularly critical of the humanitarian leadership role within the UN, but it seems inexplicable that the absence of UNOCHA, the UN agency with