

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 paper¹ for the ALNAP October 2000 symposium². 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 is 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.

The majority 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: 'Though 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 practice³, 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 consultancy tends 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 to 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 problems 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 on 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: 'The 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 that 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 from the viewpoint 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 takes the diametrically opposite view that its onset was far from sudden and was clearly to be foreseen 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 this role 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. The analysis of civil–military relations is limited primarily to NATO, however, 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 mentions, let alone considers, the different meanings of ethnicity (or 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 work⁴ 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 for the writing of terms of reference.

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: 'During 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. This set of evaluations is no exception. While several note that the overall humanitarian response in Kosovo was driven by geopolitics (pre-dominantly 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 capacity for rapid emergency response. 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 special mandate for humanitarian coordination, attracts absolutely no attention in the evaluations.

UNHCR, as designated UN Lead Agency, serves throughout as chief whipping boy, with the lowest marks for achievement (for example, due to its 'missing middle management'), although higher marks for effort and achievement as events unfolded. WFP gets the top mark for effective deployment of human resources and logistics. UNOCHA, confined to arranging military and civil defence packages from Geneva, producing situation reports in Macedonia, and coordinating with the Albanian government's EMG (Emergency Management Group), was virtually absent, so goes unmarked. UNHCR rather than UNOCHA participated in key EU-NATO coordination meetings.

Lack of strategic planning, and not just information sharing (considered by some evaluations to have been reasonable), carries a share of the blame. This lack was highlighted by Dr Bernard Kouchner (UN Secretary General Special Representative) at an August 1999 meeting in Pristina, when he lamented the fact that there was 'simply no strategic intelligence available' to allow him to tell a visiting new donor (Taiwan) how best to support the relief effort.

The imbalance of participation in Kosovo from within the UN system, as well as contributing to the overall coordination crisis, may reflect a coordination crisis within the UN itself.

A comparison of the imbalance of participation of big UN players with non-UN (but still big) players, including ECHO and OSCE, would be an interesting evaluation focus, as would a focus on UNDP (another part of the UN with notable coordination capacity and experience). It is likely, however, that only a system-wide evaluation would attempt this.

3.2.8 A crisis of proliferation

As with Rwanda, so with Kosovo, what all commentators call 'an unusually large number' of NGOs responded to the crisis. One document famously speaks of this conflict having attracted one NGO for every day of the year. Unfortunately no report goes beyond this rather simplistic 'large number' plaint to tell us just what all these small (and smallest) organisations do, where they come from, what resources they have, etc. They only reveal that having them there makes coordination difficult or nigh on impossible. Their message is of the 'once again' variety where lessons have not been learned.

On the face of it, this is an extraordinarily limited, unconvincing, and trivial assessment. It is most unlikely that fewer of the smaller and smallest would have made that much difference to balanced participation and coordination. It reads less as evaluation than as special pleading.

3.2.9 A crisis of unpredicted and unpredictable scale and speed

Many evaluations share the ‘unpredictable nature of the crisis’ view. First of all hundreds of thousands of displaced people streamed out of Kosovo and then, equally rapidly and unexpectedly – perhaps especially from Macedonia but also Albania – they streamed back again. To whom was this unexpected?

The evaluations merely report that this was not expected by the international relief agencies. Both the evaluations and the more research-based policy reviews are completely silent on the speed and scale of movement as seen and managed by the affected populations themselves. This is another instance of the virtual reality of ‘foreign-aid land’, owing to its social blinkering (and prejudice) about the social, cultural and other institutions of the populations it is meant to be serving.

3.2.10 A host nation crisis: maintaining domestic sociopolitical stability

In this complex emergency, governments were in place in Macedonia and Albania (as of course in Serbia). The Macedonian government in particular considered that the refugee outflow from Kosovo threatened its stability. At the peak of the crisis refugees from the Kosovo conflict constituted 15% of the Macedonian population.

The threat to host nations’ domestic sociopolitical stability is, however, only dimly seen and appreciated by the evaluations, where the emphasis in this respect is always on humanitarian and human rights principles and international refugee law issues. Yet, if the Macedonian government had fallen this would undoubtedly have made the whole humanitarian scene even worse.

Instead of treating issues of regional government stability as integral to the humanitarian context, the tendency is to leave the former completely outside the latter’s framework, benefiting neither in practical nor policy terms.

What programmes and their evaluations do seem to have been aware of, however, was the regional poverty factor: ‘The countries that bore the brunt of the Kosovo crisis are among the poorest in Europe,’ (DEC Vol. I, 2000c: p xvii); ‘MCIC [Macedonian Centre for International Co-operation] in Macedonia explicitly voiced the need to “balance the refugee crisis with the social crisis” ’ (DEC Vol. I, 2000c: p 45).

3.2.11 A crisis of overbearing humanitarian assistance

Fortunately, two or three of the evaluative reports make reference to additional problems created by the humanitarian response process itself: ‘[An] agency complained that the population was fully employed in reconstruction until the international community intervened,’ (DEC Vol. II, 2000c: p96). Dispossession is common in foreign-aid land, however well intentioned.

The dangers of negative impact are not unrecognised, but warning voices clearly went unheeded. The following was written in April 1999: ‘For a decade, ethnic Albanians have been forced to create their own parallel society, involving every domain from kindergartens to health clinics to architecture schools to a vibrant independent media. It is critical that the international community not “colonise” Kosovo with a sudden “invasion” of well-intended but overbearing humanitarian assistance. Such a misguided deluge of aid last autumn, in the wake of the ill-fated October [1998] ceasefire agreements,

seriously jeopardised the integrity and growth of Kosovo's home-grown, still fragile, civil society institutions. It is clear from the many mistakes made by the international community in its efforts to rebuild Bosnia that a peremptory, non consultative approach using "in-and-out" commercial contractors can inhibit the efforts of a vulnerable, war-torn, population to get back on its feet,' (International Crisis Group, 1999a: p27).

3.2.12 A crisis in assistance: assisting a 'non-crisis'

Despite all of the above, a theme that appears in some of the evaluations (particularly with reference to assistance) is that 'no classically defined emergency actually occurred,' (UNICEF/DFID, 2000: p13); 'The refugees were relatively better nourished, healthier, and with access to greater resources (including, very significantly, from the Kosovar Albanian Diaspora) compared to those in many emergencies ... [and] mortality and morbidity rates were generally well within emergency levels,' (ibid.). Also: 'The humanitarian intervention as a whole can only take very limited credit for the low mortality and morbidity [noted by many commentators]. The quality of the sanitation and environmental health interventions would probably not have been sufficient to prevent large-scale breakdowns of public health, epidemics and, as a result, excess loss of life, had the crisis occurred in a more usual emergency context,' (DEC Vol. II, 2000c: p25).

WFP's recognition that 'coping mechanisms available to many of the affected groups ... played a decisive role in avoiding hunger and malnutrition ... [as] refugees, IDPs, and other vulnerable groups had their own contingency plans, emergency food stocks and resources to overcome periods of dislocation from their homes ... [and] themselves assumed responsibility for their survival and well being, particularly during the first weeks of the emergency,' (WFP Executive Board, 2000a: p5) should also be carefully noted.

However, no organisation is reported as having considered there was anything but a huge programme of assistance to be undertaken, or that such efforts were inadequate or, in some regards at least, unnecessary.

3.2.13 Calling a conflict a crisis: some discourse analysis

A final important point is that most of the evaluative reports, particularly the evaluations, tend to speak not of the Kosovo 'conflict' but of the Kosovo 'crisis' or 'crises'.

Talk of 'crisis' tends mainly to ask 'what?' questions (for example, What is the crisis? What is the assistance required?). 'Conflict' talk triggers more 'who?' questions (for example, Who is responsible for this mess? Who is fighting whom?). Of course there is no watertight divide between 'crisis' and 'conflict', as the words have overlapping as well as different dimensions.

The word 'conflict' makes only very rare appearances – generally where some broader background or scene-setting chronology is done. While the humanitarian response in Kosovo is not charged with fuelling the fire, the evaluations do say that much of the response missed the fire, in respect of protection and coverage. At best, only a third of those affected (those in refugee camps) received assistance.

The discursive practice of calling a conflict a crisis is not however without effect, intended or not. It focuses problem identification on crisis management rather than conflict resolution, peace and development, de-politicising the analysis. This leads to organisations such as the Mother Teresa Society and the Kosovo Liberation Army (KLA), both prominent local players, being in effect written out of the humanitarian scene and its analysis. They are labelled 'political' or 'military' and hence not

‘approved’ civil-society organs, which humanitarian agency discourse takes to be apolitical, somehow beyond politics and war. It is a grave gap in evaluation.

Not one of the evaluations addresses its own, or its programme’s, mode of discourse. Commissioned evaluation as a genre is rarely self-reflective.

3.2.14 **Conclusion: whose crisis anyway?**

The chief findings of this section are that the various actors were responding not to one crisis but to many crises, and that differences in perception as to the nature of these crises led to the use of different success/failure indicators.

The question of whether the crises were found, perceived or made was not asked by either the evaluations or the policy reviews, explaining their tendency to dwell on and over-emphasise donor relief supply-driven factors.

This myriad of crises (perceived or not) will of course have implications for whoever may take on the task of commissioning and planning a system-wide evaluation. It will need to look not only at all the players (international or national) but also at all the crises that lie within this single conflict.

3.3 Agency Response Issues

3.3.1 Introduction

This section provides some synthesis of the evaluations' principal findings, conclusions and recommendations – with elements of meta-evaluation. They fall into three broad categories:

1. humanitarian policy, practice and strategic planning for emergency response;
2. organisation and management (O&M) issues for emergency response;
3. technical and standards aspects of emergency operations.

3.3.2 Humanitarian policy, practice and strategic planning

To prepare or not to prepare?

All the evaluations consider preparedness as an issue and most take the view that none of the agencies involved seriously anticipated the scale and speed of the emergency. They find agencies' preparedness capacities greatly in need of improvement, but few make even minimal proposals as to how this might be achieved. There is no examination of how the preparedness lessons should be incorporated into policy, practice and strategic planning.

Given the poor record of early warning systems, it is suggested that investment should be made in the development of rapid reaction rescue capacities: 'The failure of "early warning" in the Kosovo case confirms the historic tendency of such systems to be unreliable or inadequate. Rather than develop its "early warning" capacity, UNHCR should strengthen its mechanisms to react rapidly,' (UNHCR, 2000a: p xii).

Conflicts of principle: neutral versus impartial assistance?

It is important to understand the difference between neutrality and impartiality. The principle of 'neutrality' being that of not taking sides in hostilities, and the principle of 'impartiality' that of non-discrimination, a commitment to deliver relief solely on the basis of priority-of-need, regardless of allegiance.

The evaluations of and for INGOs discuss at length whether the military can pursue military and humanitarian objectives that are both independent and simultaneous in a principled humanitarian way. Most of the evaluations conclude that it cannot, seeing a fundamental contradiction in principle. Other reports, particularly those from NATO's lesson-learning workshops, take an opposing view.

Conflicts of principle between military and civil agencies are exemplified by one of the more policy-oriented studies. The issue it addressed most directly was 'whether the Kosovo crisis exemplified the harnessing of the military for humanitarian tasks or the militarisation of humanitarian action, or some combination of both,' (Minear, 2000: p viii). Although many of those involved in the Kosovo crisis, both military and humanitarian contingents, saw this as 'the harnessing of the military for humanitarian tasks. Yes, the harness chafed here and there, but on balance the collaboration was productive. ... we as researchers read the Kosovo crisis as reflecting the militarisation of humanitarian action, with certain ominous portents for the future particularly in terms of the politicisation of humanitarian access and activities,' (Minear, 2000: p viii).

The differing views are characteristic of the corpus of Kosovo reports, and the issues introduced or intensified by NATO's involvement are seen as vitally important and underlined by all. But while some are afraid it presages 'humanitarianism by default' because of military mandates, objectives and

conditionalities, others are more content to emphasise the immediate and temporal specificities, hoping that NATO's engagement brings only limited implications.

Areas of possible conflict of principle and image between military and civil agencies are only part of the picture. Another perspective on the 'conflict of principle' in the Kosovo context is the perception held by many within western society, as well as other cultural traditions, of humanitarian action as a new form of western imperialism.

Finally, clashes of commercial, along with civil and military, principles and practice need to be considered as well: '[T]he humanitarian marketplace. It becomes increasingly competitive, not least within the field of emergency response, and increasingly driven by donors and their priorities,' (DRC, 1999: p20).

Practice ahead of principle: humanitarian evacuation and humanitarian transfer

The particular pressures of the Kosovo conflict, fast and massive outflows of refugees and resistance to hosting by regional governments combined with media demands, led to the introduction of two new strategies by UNHCR. These were Humanitarian Evacuation Programmes (HEP), involving movement out of the region, and Humanitarian Transfer Programmes (HTP), involving in-region cross-border movement.

These are examined in three or four of the evaluations with, on the whole, similar conclusions. Success in addressing such massive refugee flows would have been unlikely if such innovative practices had not been employed – i.e., principle needed to catch up with practice. 'The innovative HEP resulted in an operation of unprecedented speed and scale that contributed positively to the protection of refugees by alleviating the burden on a reluctant host state that feared destabilisation,' (UNHCR, 2000a: p102).

Problems associated with both are explored including issues of design, execution, vulnerability, eligibility and fairness in selection. '... HTP did not contribute significantly to protection during the emergency. Part of the limited implementation of HTP relates to confusion over the degree of voluntariness or consent required for these refugee movements ... HEP undermined HTP: UNHCR's stated preference for protection options within the region becomes difficult to maintain if evacuations outside the region are available and preferred by the refugees,' (UNHCR, 2000a: p102).

The non-camped refugee population

A substantial proportion of the total refugee caseload found shelter in what the humanitarian community first called 'host families' and later 'private accommodation'. This latter reference is perhaps not only because payment passed hands but also because non kith-and-kin relations were involved. This is not explored in the reports.

Although they represent the majority, little is revealed about the assistance and protection provided to the Kosovar refugees in Macedonia and Albania that did not go to camps. In Macedonia local religious charities and governments, rather than the international community, probably deserve the principal credit for what was made available. Needs assessment of these non-camped affected populations was even more neglected than that of those in camps, though arguably: 'It was not significantly relevant to the well-being of the majority of the refugee population who were in private accommodation,' (DFID, 2000: p5). However, another notes: 'For the first two months not a single food distribution reached the 27,000 refugees in private accommodation. ... This failure to reach refugees living with host families

will be looked upon (alongside the failure to complete the registration of refugees while they were in Albania) as among the gravest shortcomings of the aid effort in Albania,' (Porter, 1999: p22).

The term 'vulnerability' also took on a special sense in this complex emergency. Given the extent to which refugees were in receipt of remittances from relatives living and working, for instance, in Germany, it referred not, for example, to the elderly or the sick, but to those who weren't in receipt of such remittances. There is some evidence that only the poorest went to the camps and that some 'host families' charged rent for the 'private accommodation' they provided. According to Australian media reports the evacuees airlifted to Australia complained that their temporary (Australian) accommodation was far from consistent with what they had had at home.

Coordination

As mentioned in the previous section, what emerges from the evaluations and other evaluative reports is the complexity of the coordination problem. Individual evaluations tend not to take a comprehensive view but focus only on one aspect or another, making pronouncements accordingly. Furthermore, different stages in the conflict require different assessments. Overall, accounts agree that in certain areas there was little policy coordination to begin with, though this became marginally better as the emergency wore on. UN focal points such as food (WFP) and aspects of education (UNICEF) were apparently very effective throughout and it is also reported that, within and across some more specifically INGO concerns, cooperation occasionally went well.

What varies from one programme or agency to another, in respect of what was well done, may depend on the extent to which each consciously built on earlier experience and evaluations of that experience. The reports however give greater attention to explanations for lack of achievement: 'For reasons that are well-known, the Kosovo crisis brought in very substantial funds from member states and each wished to maintain a high profile

in-country, including undertaking their own programmes. This led to considerable confusion and lack of coordinated activities,' (ECHO, 2000k, 2000l, 2000m: p5). But note also that: 'Within a few weeks as donor liaison mechanisms also gained in effectiveness, a more coherent and consistent needs-driven approach was evident. Coordination, nevertheless, proved particularly difficult to handle in this emergency. ... We should have articulated the many problems and constraints more clearly from the start,' (Morris, 1999: p17).

What may come as a greater surprise is the comment that, despite shortcomings, adverse effects can be exaggerated: 'UNHCR shortcomings [in assistance and coordination functions] ... did not have grave consequences for the welfare of the refugees: indeed they were relatively minor in relation to the overall relief response. ... [Nonetheless] areas of demonstrated weakness and inability to rapidly meet its own standards of response affected the credibility of the agency [which therefore could be considered to have suffered the most]', (UNHCR, 2000a: p xi).

While the evaluations mention the constraints of consensual coordination, none make recommendations for change in this area nor examine it in other ways.

What constitutes an emergency?

As with any disaster or humanitarian crisis that occurs in a sovereign state, a state of emergency cannot be acted on until it is declared and recognised as such by the national government concerned. The simple question of whether there is an emergency to meet may be part of the reason for little or no

response at the outset. It would also appear that key institutions lack clear rules and guidelines for consideration of what constitutes an emergency status and how it should be acted on.

Where guidelines exist, to what extent are they known and used? ‘The underuse in UNICEF (even ignorance) of the emergency handbook throughout the operation is notable,’ (UNICEF/DFID, 2000: p34). The same evaluation comments: ‘Software may not be totally emergency friendly,’ (ibid.: p34).

In terms of whether a situation merits being regarded as having emergency status or not, one evaluation reads as follows: ‘Between April and May 1999, morbidity and mortality rates in Albania and Macedonia remained below the threshold that signifies an emergency. ... [F]ar below the [mortality] threshold of greater than 1[death]/10,000[people]/day, the key Sphere indicator for an emergency. ... Data for this period are not available from Kosovo or the rest of the Balkans. Since the return of the refugees to Kosovo a comprehensive health study undertaken among the Kosovar population indicates similar low levels,’ (DEC Vol. II, 2000c: p21).

There are also different types, scales, and durations of emergencies, each perhaps with their different requirements. With UNHCR the problem was partly that it ‘responded to the Kosovo refugee crisis as if it were a “normal” emergency. Standard routines for a smaller or slower emergency were followed (although not always attained),’ (UNHCR, 2000a: p x).

It is to their credit that some of the sources draw attention to these important considerations. Training issues that need to be informed by them are mentioned under the heading ‘Human resource weaknesses’ (see [3.3.3](#) below).

A further area of concern is the use of indicators. Normally in evaluation a different concern is considered to deserve a different indicator, as is the case in most of the Kosovo reports. However, to gauge the success (or failure) of a programme in terms of, say, a ‘no one died of famine or disease’ claim, when neither famine nor disease constituted the emergency is, as some evaluations say, downright misleading, if not aberrant.

3.3.3 Organisation and management of human resources

Lack of standing rapid response capacity

As mentioned earlier, another striking story to emerge is that of an international humanitarian community that lacks a rapid response capacity. Rescue here does not refer to the humanitarian evacuation programme that eventually got underway in Kosovo, but to a strategic concept for initial agency reconnaissance. The US government’s DART (Disaster Assistance Response Team) approach, which emphasises the assessment required for such rescue, may come close to what the international humanitarian system lacks.

Most evaluative reports, including those by the military, lament this woefully inadequate situation. For example: ‘Recommendation 1: If ACT is to be operational [in emergencies] it must have an emergency management team that it can deploy immediately to make preliminary assessments and to prepare programmes. Where a multi-agency response is involved, this should include preparing common programming,’ (ACT, 2000: p10).

The lack of a specialist standby rapid response, whatever may account for it, is an absurdity. If one is ever established, particular attention should be paid to the rapid assessment skills required and whose capacities are to be built – those of ‘foreigners’, or of ‘nationals’. What are needed in the first instance

are rapid response assessment teams, not rosters of active-service personnel. One evaluation notes that emergency response capacities have yet to be mainstreamed into UNICEF.

Although the above is illustrative of the wider picture, it is encouraging to note that in response to a recommendation put forward by the evaluation WFP commissioned, the Emergency Preparedness and Response Steering Committee was created at executive staff level to ‘establish a better framework for early warning, preparedness and response to rapidly evolving emergencies,’ (WFP Executive Board, 2000b: p1).

UNHCR’s evaluation noted that UNHCR was ‘not established to provide relief in emergencies ... If members of the General Assembly want to do this, they will get a refugee agency that is quite different from that they established 50 years ago.’ It noted, however, that: ‘It is clearly possible to turn UNHCR into a superbly efficient rescue service,’ (UNHCR, 2000a: p xiv), if and when this is the policy decided.

Training and briefing

Several evaluations comment on the paucity of available training material and consider this to be partly responsible for poor practice. This is another example of a crisis of the agencies’ own making. Happily some evaluation terms of reference called for an assessment of personnel orientation, guidance and training before and during deployment, but the evaluations generally paint a very gloomy picture. When does poor practice qualify as malpractice?

In terms of briefing: ‘[S]ome pointed out that if they had been briefed about field work and the importance of cooperation ... by a veteran field worker, the message might have been more convincing than coming from a desk worker,’ (DRC, 1999: p25). Debriefing is also crucial, as is the use of the information gathered from the undertaking.

Agencies were noted as having recognised the useful combination of in-house training with external courses by organisations such as RedR, Sphere, and Merlin, but the reports, despite their vocal criticism, offered little remedy other than one interesting suggestion that a travelling contingent of field-craft trainers should be established.

The symposium background paper, on which this chapter is based, proposed something akin to the ‘learning support office’⁵ currently being developed by ALNAP.

Human resource weaknesses

Poor programme performance can often be attributed to poor human resource management (HRM). Unprofessional employment and people management appears to have been much as usual in this complex emergency. A familiar pattern emerges in respect of short-term assignments, lack of appropriate training, even lack of briefing on arrival. Gender is reported to have been mostly ‘forgotten again’, and the high level of staff turnover a destabilising factor, particularly where agencies lack systems and procedures to ensure continuity and enable new staff to become effective more rapidly.

Shortcomings are also identified in relations between foreign and national staff; between country offices and points of delivery; in the mismatch of specialists and generalists (mostly the absence of the former where required sectorally); between routine and volunteer personnel; and in respect of conditions of service and duration of deployment/assignment. One evaluation notes that: ‘[P]eople

who have been deemed unfit for previous operations in the Balkans appeared again in the Kosovo operation,' (DRC, 1999: p25).

Weaknesses in the management of organisations, especially middle management, are easy to diagnose. A common finding in many of the evaluations is that: 'In massive emergencies the agency should ensure the rapid release of middle managers by the immediate adoption of directive, rather than voluntary, deployment practice,' (UNHCR, 2000a: p xvi).

These weaknesses are responsible for many of the undesirable actions and outcomes in the programmes performed. However, other than in particular cases (primarily the O&M reports), little provision is made either in the terms of reference or the evaluation team composition to allow organisational and managerial dimensions to be competently addressed. This is probably why there continues to be so little progress in the evaluation of these areas, even though the evaluations make a few more proposals on the remedy for this than they do for training and briefing.

All the evaluations portray Kosovo as yet another case where a large-scale humanitarian response pressured agencies into recruiting and deploying personnel without training, although UNHCR personnel did report that they had received a certain amount of appropriate training.

The problem seems to lie particularly within INGOs, where even senior staff members were found to be unaware of their organisation's commitments as signatories to codes that affirm, among other things, certain human resource management norms: 'Although 5 of the 12 DEC agencies are signed up to the pilot People in Aid "Code of Best Practice" some senior staff in DEC agencies were unsure whether their agency was a signatory or not. It is not possible to say how well agencies conformed to the code,' (DEC Vol. I, 2000c: p xxvii).

The importance of existing presence and national staff

It appears to be the exception rather than the rule that agencies insist on existing presence and national staff as a condition of response. However, in addition to improved timeliness and contextual understanding, it is suggested that agencies working through local partners are less likely to pursue their own agendas.

Two evaluations clearly note existing presence and national staff as enhancing factors: '[T]he effectiveness of this church work is due to a combination of factors: 1) they already operate on the ground ... ; 2) they had staff in place (whether national or missionary) who spoke the local language ...' (Tearfund, 2000: p27); and in the case of UNICEF: 'The predominance of national staff in middle management positions had a direct impact, in that it facilitated a deeper understanding of needs and realities in the three countries visited than is the case for international agencies managed almost uniquely by internationals, most of whom do not speak the relevant languages,' (UNICEF/DFID, 2000: p5).

Lack of social knowledge and learning

Another important aspect of the story to emerge from most evaluations is the 'once again' factor. Once again 'foreign' humanitarian workers arrived ill-equipped in terms of their sociopolitical and cultural knowledge of the conflict-affected populations. Despite the code of conduct that the big INGOs have all signed up to (introducing an ethical dimension), social and cultural ignorance of beneficiaries was the norm. The heavily loaded term 'beneficiaries', as discourse analysis would say, accounts for a large part of the problem.

An obvious example is the criticised UNICEF ‘Superman Project’ developed as part of a mine-awareness campaign. As well as conveying the wrong message – that Superman will save you – when the intention was aimed at inculcating safer behaviour: ‘Handicap International psychologists recommend the identification of role models within the community whose behaviour children can copy, rather than the use of fictional heroes ... [T]he age range of the target audience (7–14 years) is too wide, there are no cluster bombs featured although cluster bombs form part of the threat, and the material is in any case not immediately available in Serbian,’ (UNICEF/DFID, 2000: p24). This highly illustrative example is typical of many from other agencies.

Since it is well recognised that people in the face of a crisis deploy a range of coping strategies, including the decision to move, why did the programmes and their evaluations discover so little about these strategies? One factor is that results-based management is neither oriented nor sympathetic to social knowledge needs, another is the genuine difficulty of systematic consultation in the first phase of an emergency response. This is, however, no excuse for a lack of consultation in evaluations that come later with more time.

The majority of evaluations do too little to cross the knowledge divide but the very nature of complex emergencies is that they are just that, complex: ‘Consultations and cooperation with local communities are part of DRC [Danish Refugee Council] project practice. But emergencies, it was argued, left no time for consultation other than those with local authorities. The principle, all agreed, was valid. But the question was how [to do so] in the heat of the emergency,’ (DRC, 1999: p22). Agencies’ basic day-to-day work inevitably involves adjustments and learning, clearly not undertaken in a social vacuum.

Even where evaluation time is dedicated to learning about beneficiaries, through in-depth individual and family interviews, the focus is mainly on discovering whether assistance provided was appropriate rather than learning, for example, why a return ‘more spontaneous than organised, and self-resourced than assisted,’ happened (UNICEF/DFID, 2000: p32).

Lack of organisational learning

The question of organisational learning relates to much of the above. Like most humanitarian evaluations before them, the Kosovo reports state that past supposedly learned lessons still have not been learned (or at least where learned have not been acted on). One evaluation records a member of staff asking: ‘Why are questions about contingency planning and related staffing policy only now discussed. Since many years WFP has been handling emergency operations all around the world, and WFP’s budget has shifted from the original 80% for development to almost 80% for emergency programmes. Why are constraints like these raised only now?’ (WFP, 2000b: p27).

The fact that the evaluations make many explicit comparisons with earlier complex emergencies might help speed along the information base required for such learning. However, where previous evaluations are the source of such wisdom, use of them should include critical – and transparent – examination of their methodology.

While the need to focus seriously on the process of organisational learning is not being addressed, there are signs of more robust attempts at lesson-learning from Kosovo. Evaluation terms of reference and team composition should be revisited to acknowledge the linkage between lesson-learning and accountability, rather than adopting the view that addressing both represents overload.

The UN Lead Agency role

UNHCR was selected as UN Lead Agency in Kosovo due to its refugee protection mandate under international law, although it is not a humanitarian relief organisation by mandate or, arguably, capacity. Why 'refugee crisis' should trump 'food crisis' or 'integrated and system management crisis' (and, by extension, WFP and UNDP) is not explored, but the evaluations do note a lack of conceptual clarity around the 'Lead Agency' status, and its voluntary (consensual) nature. None however discusses it seriously nor makes recommendations.

Accreditation, particularly of the large numbers of NGOs, appears not to be part of the process of participation. Although raised as an issue, no evaluation offers recommendations, and overall the evaluations pay insufficient attention to the Lead Agency as an issue.

Delivery, distribution and impact

All the reports agree that, for whatever reason (they never address why): 'Uniform tracking, monitoring and evaluation systems and approaches were not applied in the emergency,' (UNICEF/DFID, 2000: p7). Other similar statements: 'Neither the ICRC nor the Federation has a monitoring system that provides sufficient feedback,' (ICRC/IFRC, 2000: p ii); '[O]f particular concern is the absence of a system which reconciles logistics statistics (what was sent out) with relief statistics (what was distributed to beneficiaries),' (ibid.: p iii).

According to one of the military reviews: 'The US ambassador in Albania did not have a good understanding of the [Marine Air Ground Task Force's humanitarian] assessment team's role. ... The consequent lack of coordination and cooperation resulted in the departure of the assessment team before it had completed its mission,' (US Department of Defence, 2000: p106). Although UNHCR's Village Needs Assessment Unit in Pristina was seemingly much appreciated it is not assessed in any of the evaluations.

One reason why programmes do not monitor their work in this area may be that agencies still live in the charity paradigm and are simply unwilling to dedicate funds to this purpose. This sits particularly badly in the case of well or over-funded programmes.

Most evaluations that mention 'needs assessment' conclude that remarkably little was done. The extent to which, in this emergency as in so many others, relief supply-side considerations and blueprint approaches ruled, was also noted. This was exemplified by responses that at one point threatened a mini-crisis in assistance because they were so inappropriate to needs: 'Unsolicited "supply driven" in-kind donations clogged the system (second-hand clothes). In some cases they contravened standards and guidelines for emergency aid (milk formula),' (UNICEF/DFID, 2000: p12).

The following appears typical of the scene as a whole: 'Uniform tracking, monitoring and evaluation systems and approaches were not applied in the emergency, nor was adequate capacity available for these functions,' (UNICEF/DFID, 2000: p7).

Similarly: 'There is an absence of meaningful reporting. Copious information – like the volume of inputs distributed – is available on the result of projects, but [there is] a paucity about the impact on the beneficiaries along with their families and communities, limiting to a great extent possibilities for more pro-active planning in projects,' (ECHO, 2000k, 2000l, 2000m: p4).

As usual, cultural factors as they relate to needs appear to have been ignored, despite cultural sensitivity supposedly having been a factor in some recruitment. For example, most of the detailed accounts of psychosocial assistance observe that this to have been the area of service most dogged by definitional and conceptual problems, as well as by culturally inappropriate 'Western' ways of

diagnosis and treatment. However: ‘The psychosocial programme was appreciated by a large number of recipients (“*Very useful and we needed It,*” was a typical comment), even though psychosocial support was relatively slow in coming as it was treated as part of health and not as emergency relief,’ (ICRC/IFRC, 2000: pi).

It appears that food needs in Kosovo were over-emphasised, for example, in comparison to shelter requirements, and winter needs were identified too strongly with shelter needs.

This synthesis takes the view that an unmonitored programme, besides being unworthy, is, in certain respects ‘unevaluable’, particularly now that standard evaluation of humanitarian action criteria – e.g., appropriateness, effectiveness, efficiency, connectedness, coverage, complementarity and coherence – feature widely in evaluation terms of reference.

3.3.4 **Technical and standards aspects of emergency operations**

The Sphere Technical Standards

Despite the Sphere Technical Standards becoming more widely known and discussed, with supporters seeking to expand coverage beyond the current five technical sectors, the evaluation reports (and ALNAP Kosovo Symposium plenary discussions) indicate that they remain controversial, both in respect of their value in principle and their applicability in practice, where operational flexibilities and strategic priorities must carry the day. The principal findings of the evaluations as a whole are summed up in the following: ‘The evaluation team found that awareness and application ... of the Sphere standards was poor within most DEC agencies, their international networks and local partners. ... Some DEC agency staff questioned the relevance of [the] Sphere standards in a European setting ... [H]owever, the evaluation team felt that most of the Sphere standards were applicable, particularly those relating to assessment, monitoring and information systems,’ (DEC Vol. I, 2000c: p xxvii).

Codes of conduct and operating principles

As already noted, evaluations found little evidence that codes of conduct were familiar to the majority of agency or implementing partner personnel: ‘The evaluation team found that awareness of the code of conduct and its principles ... was poor and little attention given to their application. ... Many local partners had never heard of either the code or Sphere standards, and where DEC agencies worked through local or international networks, other agencies in the network did not necessarily share the same commitment to standards. There was very little monitoring of adherence to the Code and standards in agencies’ planning and implementation of programmes and no requirement from the DEC to report against them,’ (DEC Vol. I, 2000c: p75).

3.3.5 **Conclusion: saving graces**

The findings, conclusions and recommendations above amount to what can only be called a severe critique of the aspects of the programmes they address. It is a damning indictment of the humanitarian system that the findings and recommendations synthesised here are neither new, nor Kosovo-specific.

That the less defensible aspects of the international response (including failure to reach the majority of those presumably in need) did not have a greater negative impact owes more to luck than judgement. It deserves to be remembered that, in this instance, needs were reported as apparently lower than typical in complex emergencies.

The expression 'saving grace', used in one evaluation, is taken up in another: 'The independent evaluation of UNHCR performance during the Kosovo refugee crisis acknowledges that the two main "saving graces" were the hosting by families and the refugees' ability to pay for rent and food. ... It also notes that this situation was unsustainable. In other words, the third "saving grace" was the short duration of the crisis,' (DEC Vol. II, 2000c: p67).

This synthesis adds three saving graces of its own: the stability of the governments in the region; NATO's humanitarian contribution (though NATO is not normally considered part of the international humanitarian system); and the strength of public opinion in terms of not wanting 'another Bosnia'.

Almost without exception, the evaluations fail to explain what is arguably the most prominent social feature of this conflict and its response. That is, how the affected populations themselves saw and managed both the conflict and the vast and rapid movements that constituted their response to it. As well as elements of spontaneity and reflex, there must have been some institutional, structural and social capacity factors.

This synthesis cannot treat the issue substantively due to the absence of social information and analysis in the reports considered. As a result, it is hard to reconcile, for example: 'The generous mentioned provision of more or less timely and adequate food to the target groups by the WFP and other major food providers,' (WFP, 2000b: p20); 'mortality and morbidity rates were generally well within emergency levels,' (UNICEF/DFID, 2000: p13); and, 'the finding that the majority of refugees and internally displaced persons were in fact well nourished meant that there was little need to distribute the high protein biscuits,' (ibid.: p19).

That a good part of the story is, anyway, one of a human rights crisis turned into a humanitarian crisis is more to the point.

3.4 Meta-evaluation: Kosovo Evaluations as Learning Tools

3.4.1 Introduction

This last section is an exercise in meta-evaluation of the broader enterprise of evaluation of humanitarian action itself, as revealed by the reports. The aim is to determine and then deconstruct some aspects of its nature and scope as currently practised, in order to look for the comparative advantage, and disadvantage, of such an evaluation as a learning tool. The purpose is to come up with, and consider proposals for reform.

Evaluation's dual objectives, lesson-learning and accountability, may make it a less effective learning tool than activities dedicated solely to learning – e.g., lesson-learning studies or learning workshops. If so, where ought the emphasis to lie in terms of a reasonable expectation of the commissioned evaluation type? This chapter shares the concern expressed in three or four of the Kosovo studies that evaluation is perhaps currently asked to do too much, and forced to grapple with too many types of expectation, for which insufficient provision is made.

Part of the problem with present practice is that evaluations are more in line with what could be called knowledge-enhancing practices than with contributions specifically to knowledge management and implementation. As lesson-learning practice, it is not critical enough – indeed in some regards distinctly *uncritical*. This can be traced to the fact that evaluation is often restricted in its focus to the stated policies and objectives of the programmes being evaluated. Assuming these to be sound, it proceeds to verify their attainment.

Other limitations stem from the lack of serious social and political analysis apparent throughout. This leads to remarkably simplistic ideas about treating the affected populations in terms of having 'ethnic problems' only. As a result, the social, cultural and political local context of programmes almost never receives the quality and quantity of attention it ought, even in those reports that at least attempt to give it attention.

Such limitations are also clearly a manifestation of the constricting timeframes imposed on the consultancy framework. The adoption into humanitarian discourse of the phrase 'humanitarian crisis' instead of 'Kosovo conflict' has insulated the evaluative reports from even 'ethno-nationalism', a term that appears only once or twice in the 50 or so sources. There is uncritical and unanalytical reference to ethnicity, as if this were all that social and political analysis of the local and national context required.

The paucity of training is an important factor emerging as part of the story told by the evaluations, but is also an issue for evaluators in respect of the skill composition within evaluation teams.

3.4.2 Humanitarian agencies, actors and acts

All the Kosovo evaluations agree that evaluations should focus on more than just operations (and agencies). However, they are often discouraged or prevented from doing so by their commissioning organisations.

What an organisation is and what it does is never solely defined by mandate. For one thing, few such mandates are descriptive, rather, they are exhortatory. For another, organisations have staff, resources,

histories, problems, strategic objectives and so on, each an influencing factor in what an organisation is and does. This meta-evaluation takes the view that other levels of analysis are equally important, in particular the analysis of individual actors and their acts.

As shown earlier, individual actors (agency personnel) may not be familiar with, let alone observant of, the codes to which their organisation subscribes, or even their organisation's mandates. A focus on acts allows for a comprehensive assessment, whatever the mandate and ethical code of an organisation. Even where mandates and codes are known, how they are understood and interpreted may be more influential to what is said and done than what a mandate says and intends. The evaluative reports show that beliefs vary considerably, even within the same organisation. Saying and signing is not necessarily believing and practising. The importance of circumstance and situation is above all not to be written-off in emergencies – hence the importance of the post-modern approach in this chapter.

Humanitarian organisations consider themselves to be non-commercial but nevertheless perform certain commercial acts, such as surviving as organisations through marketing and public relations. Similarly, commercial organisations (in this instance the military) have non-commercial acts to perform. In other words, not everything a relief organisation imagines, says, believes, and does, is relief and not everything the military imagine, say, believe, or do, is concerned with combat. There are multiple realities that determine what a particular organisation does in a particular situation, which may throw up concerns about the degree of 'prostitution' of humanitarianism. A debate about just humanitarian versus non-humanitarian organisations (i.e., the military) misses out too much to be practical or intellectually acceptable.

In sum, it is important for evaluations of humanitarian action not to accept stated objectives, policies and principles as givens. A more thorough and nuanced approach will be able to account for the realities of practice in relation to principle.

3.4.3 Policy evaluation contrasted with project evaluation

Evaluation of humanitarian action tends to lean heavily towards project evaluation, this being the primary reference of the increasingly standard criteria: appropriateness, efficiency, impact, coverage, connectedness and coherence (yet to be appraised for their value with regard to the evaluation of humanitarian policy).

Unlike previous evaluations, the Kosovo set of reports is notable for the extent to which the reports combine a policy and project/operations focus, undoubtedly due to the policy issues thrown up by NATO's involvement.

However, little more than introductory considerations on policy analysis were included in the 1999 OECD *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (pp 24–25). A key requirement of the evaluation profession will be to determine what the evaluation of project and policy have in common, as well as their differences.

While both policy and project (i.e., programme) evaluation requires recourse to objectives and indicators, these are not always provided by the agency project or programme under evaluation. The WFP evaluation remarks that: '[T]he difficulties in evaluating [WFP's response] can be found in the lack of specific objectives and assumptions formulated, as well as in the lack of proper indicators to measure achievement and assess the assumptions,' (WFP, 2000b: p2). A frustration echoed by others.

Induction-led evaluation expects to find the criteria it needs for its own purposes provided, as it were, on a plate. Where this is not the case, if evaluators cannot find ways to devise and provide these for

themselves, they are exposed to the charge of not getting their job done. The responsibility, however, must be shared by the agency being evaluated.

3.4.4 **Lesson-learning and lessons to be learned**

Lesson-learning is stated as a principal objective of most of the Kosovo evaluations (also most of the policy review documents and workshop reports). They return again and again to the fact that fundamental lessons for strategic planning, the organisation and management of the humanitarian response, and indeed its monitoring and evaluation – which may well have been recommended by previous evaluations – remain to be learned.

Most of the Kosovo evaluations consider that some of the shortcomings of the humanitarian response are due to *force majeure*. The circumstances of this conflict are seen to be exceptional in various ways, and therefore requiring special allowance to be made. This meta-evaluation has no general quarrel with special pleading, provided what is considered to be specific and perhaps excusable is clearly distinguished from what is not. However, what it does seek to ask is whether, in part, typical evaluation practice is as much to blame for lack of lesson-learning as are the evaluated organisations themselves.

Shortcomings in present and past evaluation practice are certainly part of the total picture to be explored. Terms of reference are typically overloaded. Divisions of labour in evaluation teams tend to be on a sectoral or functional basis only. Conclusions and recommendations of the ‘single best’ form – i.e., not menus of options with the advantages and disadvantages of each option spelt out – is the dominant practice. But on top of this comes the handicapping trend of seeing lesson-learning as simply filling knowledge deficits (a practice all too familiar in evaluation of economic development cooperation), uncomplemented by serious attention to knowledge management and implementation.

Current evaluation practice, team composition and so forth, is singularly ill-designed and ill-equipped to address the lesson-learning purpose that it sets, and has had set for it, unless it can shift its focus to organisational (institutional) learning. This learning is usually a task that agencies set for themselves, but to which evaluation could make a useful contribution. Unless the evaluation process recognises its role in relation to institutional learning, it will lose its status.

3.4.5 **Opaqueness in evaluation methodology**

Lack of methodological rigour results in impressionistic findings that are hard to defend, and undermine the reputations of humanitarian evaluations. Giving sufficient attention to methods and methodology is critical to both the quality of, and confidence in, evaluations.

Methodology (M) is sometimes neglected as an assumed ‘given’, understood by all. In reality there is no common approach or understanding, and denial of any use of methodology in the undertaking of an evaluation is not unknown. It is critical to recognise methodology’s multiple-layered nature. The following four layers are therefore suggested: M-1, M-2, M-3 and M-4.

M-1 is the layer dealt with by most of the Kosovo reports. This includes the timetable, who was where, when and why during the mission, the basis for the division of labour in the team and the time allocation in the field and other locations. Although extremely limited in analytical value, it reveals the basics of how an evaluation proceeded and whether it spent enough time in the field and HQ.

The M-2 layer should include policy analysis methods, evaluators’ involvement in finalising the terms of reference, team management issues, established networks, and how difficult relations in the field

were addressed – whether with representatives of governmental and other agencies, or with those of the client organisation.

The M-3 layer should cover conceptual and analytical or methodological frameworks for the study, including precise field methods for data collection, the criteria against which field information was assessed and judgements reached. It should also outline the extent to which evaluators felt the methodology worked and implications for findings, conclusions and recommendations.

Clarity of criteria is particularly important where the influence of team members, as optimists or pessimists, can be both subjective and influential – i.e., is it assessed as a positive that a programme achieved so much or as a negative that it failed to achieve everything. Such assessments should not be totally subjective.

Only one evaluation mentions methodology in this way: ‘Our assessment is qualitative. While a method of marking that awards points on a scale, of say 1 to 10, may be appropriate for a simple project evaluation, it is much too simple for evaluating agency operations in a complex historical context. ... The standard tools used are those of historical and public policy analysis,’ (UNHCR, 2000a: p3). But information about which tools and what they yielded, or not, compared with others, is not given, nor can it be deduced from the text of this example.

The final, M-4 layer, ought to guide the critical reader as to how the evaluators drew conclusions from findings and recommendations from conclusions. These reports do not. Given that the client (among others) is likely to judge recommendations not so much in terms of ‘true or false’ but rather ‘feasible or costly’, a methodology that is transparent would tell us the parameters used to reach such recommendations.

Despite the associated difficulties, it is all the more important in the face of complex field situations, multiple actors, lack of data, variable translation of field interviews, limited access to the field and beneficiaries, and difficulties with attribution, to give prior attention to methods and methodologies. However, the flexibility needed to respond to issues and ideas emerging from the fieldwork, and to cope with resistance met in the field, must be maintained. A two-stage fieldwork approach allowing an initial field visit or tour of the study area can help focus the development of appropriate methods and methodologies.

Short-term consultancy is unlikely ever to compare favourably with research, committees of enquiry, select committees, independent commissions and so on, in terms of its hard data findings. But evaluations are normally expected to yield *some* hard findings. One evaluation claims to believe, but unfortunately fails to demonstrate, that: ‘[W]ith regard to final results, the most important finding of the evaluation is the high level of satisfaction of beneficiaries,’ (ICRC/IFRC, 2000: p i). This level of aspiration is exceptional in evaluation, being more typical of research.

Greater transparency in addition to less overloaded terms of reference is needed. Everyone ought to develop frames and discourses that permit greater focus on conclusions and recommendations, including the provision of a menu of different scenario options in the recommendation section.

The comparative advantage of evaluation as a lesson-learning tool needs to be reconsidered from time to time. It is the functions, purposes and methods of evaluation, and not the genre itself, that is at issue – including the extent to which evaluation should be linked in practice with other approaches to learning (and participation and ownership).

3.4.6 **Independent evaluation: achieved, necessary, relevant?**

To what extent is the independence of the evaluation really necessary? And what does it seek to achieve? Independence may well not have been achieved even where evaluations claim that status.

Is independence more strategic than other considerations for credibility – such as expertise, authority, or sometimes nationality? Does it function more as a ‘credibility-placebo’ where it is no guarantor, in and of itself, of quality of output?

The reality of independence is that it can be tenuous, particularly where the like-by-like approach to commissioning extends to evaluators carrying out commissioned evaluations for past employers.

To deconstruct ‘independent’ further, and ask independent from what? – the configuration of particular disciplines, ideologies, sectors etc. should be considered, as well as personal values and premises. To whom does the evaluator report? An agency’s top management or board, a beneficiary union, an ombudsman, or another sort of watchdog? These issues are not necessarily unrelated to the choice of one consultant over another.

That the stakeholders that are taken seriously are not generally beneficiaries or their representatives, or the public that contributed money to appeals, is another important part of this picture. The stakeholder might find token reflection in evaluation team composition and the M1 features of methodology, but when even this tokenism is missing, greater concern arises in respect of what is done in the name of independent evaluation.

This chapter’s principal author believes that:

- mixed teams of insiders and outsiders are more likely to provide credible analysis and recommendations (most Kosovo evaluations fall into the category of outsiders only); and,
- in some circumstances, invited consultancy has a better chance of being read and heard by the client than an unsolicited paper in a professional journal.

A USAID ‘evaluation tip’ states: ‘If objectivity and credibility are key requirements, an external evaluation may be the appropriate choice, whereas if stakeholders’ ownership and acting on it are priorities, more collaborative or participatory approaches are usually better,’ (USAID CDIE, 1997: p3).

Given that complex emergencies inevitably involve many agencies and that interagency relations are likely to be an important factor in any humanitarian response, more interagency evaluations should be undertaken with their potential for increasing transparency – possibly a more critical characteristic than token independence.

3.4.7 **Paucity of social, cultural and political analysis**

Applying a humanitarian label should not obscure the political causes or consequences of a complex emergency. According to a recent *Disasters* journal article: ‘It is now part of received wisdom that humanitarian assistance in conflict and post-conflict situations may be ineffective or even counterproductive in the absence of an informed understanding of the broader political context in which so-called “complex political emergencies” occur,’ (Cliffe & Luckham, 2000). The reality of the majority of the Kosovo evaluations is a poverty of social, cultural and political analysis.

This virtual void has huge ramifications in terms of what is learned and not learned about the conflict and conflict-affected populations. Economic analysis of the cost-effectiveness of operations is for the

most part also missing, despite the fact that regional economic poverty is widely considered to have had an impact on every aspect of the response – including dilemmas for relief exit strategies.

Is this analytical *unintelligence* simply due to a lack of appropriate skills on the part of evaluators and programme managers? In addition to increased training, the growth of a rights-based approach to relief might have the invigorating effect of forcing evaluators to consider local people, their perspectives and contexts.

A final area of consideration, supported by some of the evaluations, are factors influencing the effectiveness of faith-based agencies' relief work as compared with non-faith based ones. A number of key dimensions should emerge from such a comparative study, though some might relate more, for example, to the issue of existing and perhaps continuing presence and national staff than to faith and ideology.

3.4.8 Programme outputs, outcomes and impacts

As could be expected from evaluations that, for the most part, display no social learning with regard to the affected populations, they offer little by way of impact analysis. Some note this themselves. Others proceed as if anecdotal evidence were enough. On the whole the evaluations do not meet their terms of reference in relation to social learning and impact assessment. However, the terms of reference that call for social learning, but do not make sufficient provision for its undertaking – in terms of time, resources and team composition – are equally to blame.

There also remains the problem of the extent to which evaluators press on regardless to their conclusions and recommendations, demonstrating a lack of self-awareness and an unwillingness to challenge the constraints placed upon them.

Spelling out conceptually the different requirements for outputs, outcomes and impact analysis, as compared with problem identification and inputs, might help remedy this situation. Here again social, economic, cultural and political intelligence and outcomes and impact analysis are closely related.

3.4.9 Attribution of evaluator bias

If the concept 'independent' is difficult to pin down, 'bias' may be even more so – especially as this could be considered to be ever present in all social and other enquiry. The issue for meta-evaluation is how best to detect and demonstrate it, so that allowance may be made for it.

One of the co-authors of the 1999 DANIDA synthesis report, '*Evaluation of Danish Humanitarian Assistance 1992–98: Volume 1*,' remarked on the extent to which he was able to trace 'evident bias'. Although the seven evaluations shared the same terms of reference and looked at aspects of the same donor programme (albeit in different countries), the six evaluation teams came from different institutions, with different histories of evaluation and composed of different sectoral bases. These characteristics shaped frames and discourses to the extent that if team names were covered up, anyone with knowledge of the evaluation community could have guessed which evaluator had carried out which evaluation.

However, the Kosovo evaluations did not have common terms of reference, and, though looking at the same conflict, were not all concerned with the same programme. It is therefore not immediately apparent how a meta-evaluation that was heavily induction-led might identify independent variables or go about the rest of the task of attributing bias.

As for bias in respect of this meta-evaluation, there is no reason to believe it is any less influenced by this author's beliefs, experiences and formation as a social anthropologist. These include strong assumptions that commissioned consultancy and freely undertaken research deliver very different products, varieties of interdisciplinary scope, verifiability of findings and conclusions. This in addition to the view that evaluations of humanitarian action are best met by outcomes-oriented work, as is the case for evaluations of economic development issues.

3.4.10 Conclusion

The task of meta-evaluation is not easy, not always welcome, and often not understood. It is not easy because of, among other things, the frequently apples and pears nature of the evaluation reports reviewed. It is not always popular because evaluators and their managers may see it as a threat to the validity of not just exemplars of a genre but the genre itself.

Evaluation has been widely advocated as a path to lesson-learning (as well as accountability), and despite the critique in this paper, the contribution to humanitarian knowledge and analysis that the Kosovo evaluations represent is itself a testament to what present practice can achieve.

The Kosovo reports are diverse. While some may be described as continuations of management by other means, others (some are mixed of course) suggest that evaluation might best serve lesson-learning through advocacy, verification, and witness (particularly where protection is their main concern), rather than through would-be management mandates, business and organisational solutions.

A standard evaluation thesaurus sums up what are particularly evident as tools of inductive argument and critique as: 'analogies, examples, counter examples, counter explanations, more often than ... exact rules ... and the statements it use[s] ... are rough guides to the truth, that is, hints and heuristics rather than exact laws. We use certain modifiers – like *prima facie*, balance of proof, and *ceteris paribus* – sometimes *probably* – to flag the qualifications involved,' (Scriven, 1991: p221).

Whether it would be well for different types of evaluation to proceed differently remains to be examined. With regard to lesson-learning, if the aim of evaluation is to do more than investigate mandates in order at least to meet the needs of results-based management concerns, it will have to find ways of becoming more outcomes-oriented than it is at present. This in turn means evaluation managers and evaluators becoming less socially blinkered or blind.

That more and more evaluations are being done and increasingly placed in the public domain – and given public consideration as in this publication – is, on any reading, excellent. What has not been explored in this chapter, however, is how the actual commissioning, designing and doing of evaluation could best be institutionalised in order for it to meet everything that is expected of it.

Endnotes

1. A much shorter, but structurally similar, form of this paper was presented and discussed at the ALNAP symposium 'Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo', which took place in Geneva from 17–18 October 2000. The title of that paper was 'Kosovo Humanitarian Programme Evaluations: Towards Synthesis, Meta-evaluations and Sixteen Propositions for Discussion'. It is now presented, in greater depth as Chapter 3. This takes into account discussions from the plenary sessions which focussed on building immediate response capacity, improving civil military cooperation, achieving

social learning, and reconceptualising humanitarian programme evaluation as critical learning practice. Evaluations not available earlier were obtained and considered.

It is important to emphasise that the purpose of this paper is to provide an overview of common (and dissenting) themes and messages, not to highlight one evaluation or study over another.

2. See Chapter 1 Endnote 2.
3. *See* (Scriven, 1991).
4. *See* (Pottier, 1999).
5. The objective of the ALNAP Learning Support Office project is to make a positive impact on the quality of emergency response in the field through the promotion and facilitation of three-way learning activities:
 - i. 'learning-in';
 - ii. 'learning-out';
 - iii. 'lateral learning'. It is proposed that a test Learning Support Office will be run in Sierra Leone in 2001.