Ethical Dimensions of Human-AI Interaction in Maritime Decision-Making: An Expert Perspective

Suhartini[1*]

[1] Maritime Institute, Sekolah Tinggi Ilmu Pelayaran Jakarta, North Jakarta, Indonesia

Keywords
human-AI interaction; maritime ethics; AI accountability; human override

Correspondence Email:
Suhartini@stipmail.ac.id

Abstract
As artificial intelligence systems progressively enter maritime decision-making environments — from AI-assisted collision avoidance recommendations on bridge displays to algorithmic cargo scheduling in port operations and predictive maintenance alerts in engine rooms — the ethical dimensions of human-AI interaction in these high-consequence professional contexts demand empirical investigation that current maritime ethics scholarship has not yet provided. This study examines the ethical dimensions of human-AI interaction in maritime decision-making through qualitative focus group discussions with 30 experts across Port and Shipping Management (PSM), Deck Nautical (DN), and Marine Engineering (ENG) program affiliations at STIP Jakarta, drawing on the expert practitioners' lived experience of automation transitions in maritime contexts. Four thematic categories were generated through systematic analysis: Conditional Trust Architecture, Accountability Asymmetry, Human Override Norms, and Generational Ethics Gradient. DN experts exhibit the strongest human-primacy ethical orientation (most pronounced endorsement of override norms and conditional trust), while PSM experts exhibit the most pragmatically accepting orientation toward AI decision support. The Generational Ethics Gradient theme documents a practitioner consensus that digital-native generations entering maritime professions will hold weaker human-primacy values than their predecessors, with expert participants identifying this generational transition as the most consequential and least institutionally prepared ethical dimension of maritime AI integration. The study contributes the first practitioner-grounded, ethics-framework-informed account of human-AI interaction values in Indonesian maritime professional contexts.

1. Introduction

The ethics of human-AI interaction has become one of the most urgently and extensively theorized domains of applied technology ethics as artificial intelligence systems move from research environments into high-consequence professional decision-making contexts where algorithmic outputs may directly influence human safety, resource allocation, and legal liability [1]. Aviation, medicine, military operations, and criminal justice have each generated specific applied ethics scholarship examining how AI integration affects professional responsibility, accountability, trust calibration, and human judgment — scholarship that, while theoretically

convergent in its identification of core ethical challenges, has not been extended to the maritime professional domain with any empirical specificity. The maritime context presents a distinctive configuration of these general ethical challenges: commercial vessels operating in international waters under IMO regulatory frameworks encompass life-safety consequences comparable to aviation, legal liability frameworks developed over centuries of admiralty law whose relationship to AI decision accountability has not been clarified, multicultural crew environments in which authority hierarchy and communication norms shape how AI recommendations are received and acted upon, and a tradition of professional autonomy in operational judgment — the Master's authority — that AI decision support systems both challenge and must accommodate.

These distinctive maritime ethical dimensions are not merely theoretical concerns for future consideration when fully autonomous vessels enter service. They are operational realities that practicing maritime professionals already navigate when they interact with ECDIS trajectory planning recommendations that they may or may not follow, with engine monitoring alert systems whose predictive maintenance suggestions they must evaluate against their own professional judgment, and with AI-assisted cargo scheduling platforms whose optimization outputs they must accept, modify, or override based on operational experience that the algorithm cannot access [2]. Understanding how maritime practitioners think about these human-AI interaction ethics — how they construct their trust in AI systems, how they conceptualize accountability when AI-assisted decisions produce adverse outcomes, and what they understand as the appropriate norms for human override of AI recommendations — is not merely an academic ethics inquiry but a professional formation question directly relevant to how maritime vocational institutions such as STIP Jakarta should prepare the next generation of maritime professionals for ethically responsible human-AI interaction in their professional contexts.

At the international policy level, the IMO's MASS regulatory framework's progressive automation degree spectrum — from MASS degree 1 (ship with automated processes and decision support) through degree 4 (fully autonomous ship) — implicitly raises but does not explicitly address the ethical questions about human-AI interaction that each automation degree level generates for the professionals whose judgment and responsibility are affected by increasing algorithmic participation in decision-making [3]. The framework specifies what levels of human control will be required at each degree but does not specify how human officers should ethically calibrate their trust in AI recommendations, what accountability frameworks should govern AI-assisted decisions that prove incorrect, or how the professional ethics of maritime judgment should evolve alongside increasing AI capability — questions that this study's expert-practitioner focus group data directly illuminates through the lived professional experience of practitioners who have already encountered analogous human-automation interaction challenges in conventional automated systems.

Three objectives guide this investigation: first, to characterize the dominant ethical orientations that STIP Jakarta's expert practitioner community applies to human-AI interaction in maritime decision-making through systematic FGD thematic analysis; second, to examine between-program variation in ethical orientation, identifying how PSM, DN, and ENG professional formation contexts generate different human-AI interaction ethics emphases; and third, to document the Generational Ethics Gradient that expert participants identify as the most consequential and least institutionally addressed ethical dimension of maritime AI integration.

1.2 Theoretical Framework

This study integrates three theoretical frameworks providing the analytical architecture for maritime human-AI interaction ethics analysis. Floridi et al.'s [4] AI4People framework — developed through multi-stakeholder consultation across industry, civil society, and government to articulate the ethical principles that responsible AI development and deployment should embody — identifies five core AI ethics principles: beneficence (AI should do good), non-maleficence (AI should not do harm), autonomy (AI should preserve human decision-making capacity), justice (AI benefits and burdens should be equitably distributed), and explicability (AI processes and outcomes should be understandable to those affected). Applied to maritime

human-AI interaction, these principles generate specific ethical questions: Does AI collision avoidance recommendation undermine the officer's development of autonomous navigational judgment (autonomy principle)? Does algorithmic maintenance recommendation create disproportionate accountability burden for engineers who follow and cannot independently verify algorithmic reasoning (justice principle)? Are AI maritime decision support systems sufficiently explicable for the maritime officers who must decide whether to follow or override them (explicability principle)?

Parasuraman et al.'s [5] framework for trust calibration in human-automation interaction provides the theoretical grounding for the Conditional Trust Architecture theme, establishing that appropriate trust in automated systems — neither overtrusting (compliance without critical evaluation) nor undertrusting (disuse despite demonstrated capability) — depends on the human operator's accurate mental model of the automation system's capability profile, failure modes, and uncertainty estimation. Applied to maritime AI systems, this framework predicts that the ethical quality of human-AI interaction depends substantially on whether maritime professionals possess the AI system knowledge necessary to calibrate their trust appropriately, a knowledge gap whose educational implications directly connect this study's ethics analysis to STIP Jakarta's AI literacy curriculum development agenda.

Jobin, Ienca, and Vayena's [6] global convergence analysis of AI ethics principles, examining the convergence and divergence across 84 national and international AI ethics frameworks, establishes that accountability — the assignment of responsibility for AI-assisted decision outcomes — consistently appears as one of the most contested and least satisfactorily resolved principles across all examined frameworks. In maritime contexts, where the Master's legal accountability for vessel safety cannot legally be delegated to an algorithm under current maritime law but where AI navigation recommendations may substantially influence the decisions for which the Master is legally accountable, the accountability principle generates a specific and practically unresolved professional ethics challenge that this study's expert FGD data directly illuminates.

2. Method

This study employed a qualitative focus group discussion design, conducted at STIP Jakarta from August 2025 to March 2026 with 30 expert practitioners representing all three academic programs [7]. Three separate FGD sessions were conducted, one per program group (10 PSM experts, 10 DN experts, 10 ENG experts), enabling both within-program ethical orientation characterization and between-program comparison. The expert population was selected rather than cadets or lecturers because this study's objectives require the kind of ethically reasoned professional reflection grounded in lived human-automation interaction experience that practitioners who have operated in maritime contexts — and directly encountered the consequences of automation failures, overreliance, and professional judgment conflicts with algorithmic recommendations — can provide that cadet and lecturer populations cannot.

The semi-structured FGD protocol addressed five ethical domains: Trust and AI Reliability (how participants assess and calibrate their trust in AI maritime systems), Accountability and Responsibility (how participants assign responsibility when AI-assisted decisions produce adverse outcomes), Human Override (participants' norms and practices regarding overriding AI recommendations), Professional Identity and AI (how AI integration affects participants' professional identity as maritime experts), and Generational Perspectives (participants' observations about how ethical orientations toward AI differ between experienced practitioners and younger colleagues or maritime students). Sessions were conducted in Bahasa Indonesia, audio-recorded with participant consent, and transcribed verbatim. Thematic analysis [8] was applied following Braun and Clarke's six-phase procedure, with cross-program comparison applied systematically. Member checking was conducted with four participants [9].

3. Results and Analysis

3.1 Thematic Overview

Table 1 presents the four thematic categories generated through systematic thematic analysis, alongside cross-program endorsement rates revealing both the universality and between-program variation of each theme across the expert practitioner community.

Table 1. *Human-AI Ethics Themes: Cross-Program FGD Endorsement (n = 30)*

Theme	PSM Experts (n=10)	DN Experts (n=10)	ENG Experts (n=10)	Total
1. Conditional Trust Architecture	9/10	10/10	9/10	28/30
2. Accountability Asymmetry	8/10	9/10	8/10	25/30
3. Human Override Norms	9/10	10/10	8/10	27/30
4. Generational Ethics Gradient	7/10	9/10	6/10	22/30

3.2 Theme 1: Conditional Trust Architecture

The Conditional Trust Architecture theme (28/30 endorsement) captures the sophisticated, context-dependent trust orientation that expert participants described applying to AI decision support systems across maritime professional contexts — an orientation that the Parasuraman et al. [5] trust calibration framework characterizes as appropriate and that this study's participants appear to have developed through direct human-automation interaction experience rather than through formal ethical training.

Expert participants consistently rejected binary trust orientations — neither unconditional trust (following AI recommendations without independent evaluation) nor universal distrust (dismissing AI outputs regardless of demonstrated performance) — in favor of conditional trust architectures that calibrate AI recommendation weight according to context-specific factors including the specific AI system's documented accuracy profile in similar conditions, the reversibility of the decision to which the AI recommendation applies, the availability of independent verification through alternative information sources, and the professional's own experiential confidence in their independent judgment for the specific decision type. "I trust the weather routing AI for passage planning across an ocean. I trust it much less when we're making a close-quarters maneuver in restricted waters. Those are different contexts requiring different trust calibration — and an officer who trusts the AI equally in both contexts is professionally dangerous regardless of which direction their trust error goes."

Between-program analysis reveals that DN experts' Conditional Trust Architecture is most elaborately differentiated and most explicitly tied to life-safety stakes, with multiple participants articulating a specific "life-safety threshold" below which they will not cede decision authority to any AI system regardless of demonstrated performance: "Below a certain consequence threshold, AI decision support is extremely valuable. At the threshold where a wrong decision kills people or destroys a ship, human judgment must remain primary — not because I'm certain I'm right, but because accountability cannot be delegated to an algorithm under current maritime law or current maritime ethics." ENG experts expressed conditional trust architectures most focused on the specific failure mode profiles of engine monitoring AI systems — endorsing AI predictive maintenance recommendations within the confidence bands that their direct system knowledge validated while explicitly rejecting recommendations falling outside empirically verified parameter ranges.

3.3 Theme 2: Accountability Asymmetry

The Accountability Asymmetry theme (25/30 endorsement) directly reflects the Jobin et al. [6] observation that accountability is the most persistently contested and least satisfactorily resolved AI ethics principle across national and international frameworks. Expert participants across all three programs described an asymmetric accountability perception: when a human maritime professional makes a decision based on their own judgment that proves incorrect, the professional liability pathway is clear, predictable, and professionally culturally accepted. When a maritime professional makes a decision based on an AI recommendation that proves incorrect, the professional liability pathway is unclear, potentially unfair, and culturally unsettled.

"If I decide to increase engine speed and something breaks, that's my decision and my responsibility. If the AI tells me to increase engine speed and I follow that recommendation and something breaks, who is responsible — me for following the recommendation, or the algorithm developer for making a bad recommendation? No one has given me a clear answer to that question, and the law hasn't either. That uncertainty makes me more cautious about AI recommendations, not because I distrust the AI, but because I distrust the accountability framework around following it." This accountability uncertainty, rather than performance skepticism about AI capability, emerged as the primary ethically grounded reason for practitioner caution about AI decision support adoption — a finding with important implications for the AI adoption literature's interpretation of practitioner resistance.

3.4 Theme 3: Human Override Norms

The Human Override Norms theme (27/30 endorsement) documents the expert community's near-unanimous consensus that human override of AI recommendations must always be available and actively maintained as a professional capability — while revealing significant within-theme disagreement about when override should be exercised. The consensus on availability (override must always be possible) contrasts with active disagreement on activation (when it should actually be used), generating a professional ethics tension whose specific character differs by program.

DN experts exhibited the strongest human override norm orientation (10/10 endorsement) and the most unambiguously specified override criteria: COLREGS compliance takes precedence over AI collision avoidance recommendation in all circumstances; Master's discretion regarding vessel safety is not delegatable to any algorithmic system regardless of IMO MASS regulatory framework evolution; and any AI recommendation that the watchkeeping officer cannot independently verify through at least one alternative source should not be acted upon in real time without Master authorization. PSM experts' override norm was formulated at the organizational rather than individual operational level: AI scheduling and logistics optimization recommendations should always be overridable by port operations management, and override authority should be clearly assigned and procedurally documented within port smart system governance frameworks.

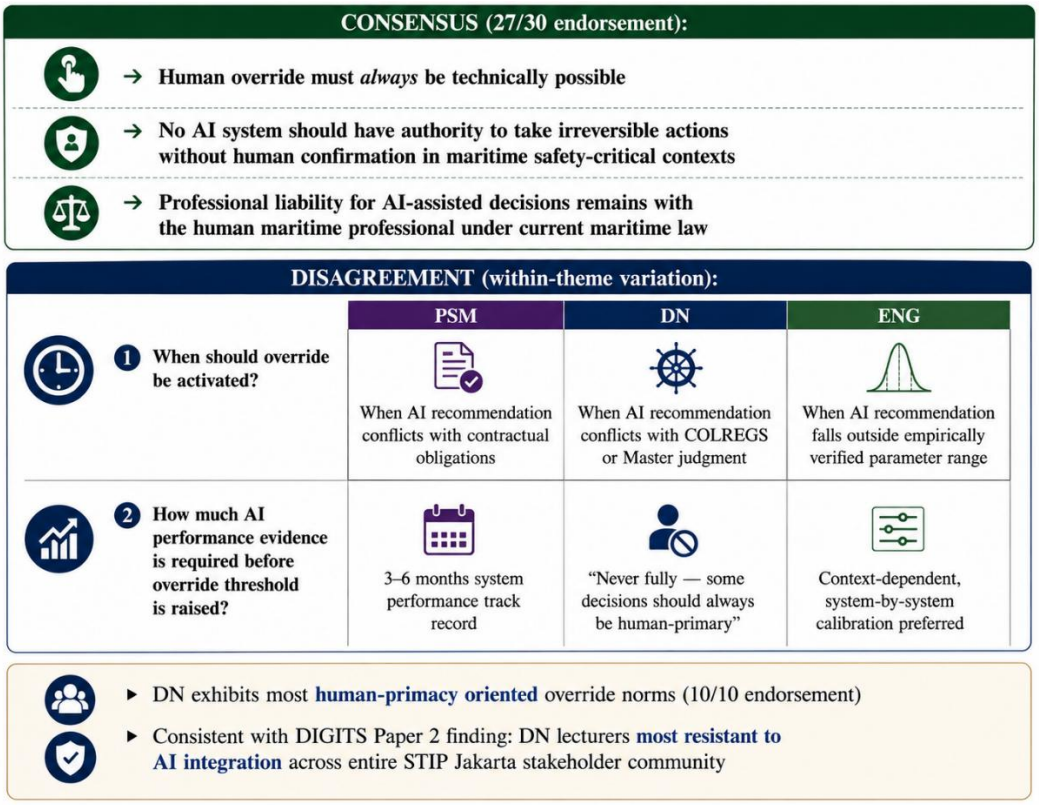

Fig 1. *Human Override Norm Consensus and Disagreement Map*

3.5 Theme 4: Generational Ethics Gradient

The Generational Ethics Gradient theme (22/30 endorsement) documents expert participants' observation — unprompted by the FGD protocol, emerging spontaneously across all three program groups — that they perceive a meaningful and potentially consequential ethical orientation difference between their own generation of maritime practitioners and the younger generation now entering maritime professions or currently enrolled in maritime education. This generational observation is particularly valuable empirically because it represents experienced maritime professionals' assessment, based on direct mentorship and professional encounter experience, of a trajectory change in maritime human-AI ethics orientation whose direction is directly relevant to maritime safety governance.

Expert participants described the generational difference not as a qualitative shift in values (younger professionals do not lack ethical orientation) but as a calibration difference in trust threshold: practitioners who trained in the era before pervasive AI assistance — who learned to navigate without trajectory prediction, plan without algorithmic optimization, and diagnose without predictive analytics — maintain what multiple participants described as an "autonomy reflex," a professionally instinctive resistance to algorithmic authority that operates even when consciously recognizing AI's demonstrated performance advantages. Younger practitioners and students, who have grown up with algorithmic assistance across personal digital life, are perceived as carrying a lower autonomy reflex — not necessarily a weaker professional ethics, but a different default trust calibration that may produce different human-AI interaction patterns in professional practice. "My students trust Google Maps more than their own spatial memory. That's not wrong for navigation to a restaurant. But I'm not sure what it means for a generation of officers trusting ECDIS trajectory recommendations more than their own seamanship judgment. It may be fine. It may not be. We genuinely don't know, and STIP Jakarta is not helping us figure it out."

3.6 Discussion

The Conditional Trust Architecture theme's near-universal expert endorsement (28/30) establishes that sophisticated, context-differentiated human-AI trust calibration is an already-practiced professional competency among experienced maritime practitioners — one that has been developed through operational experience rather than formal ethical instruction. This finding has a direct and important educational implication: the ethical competency that expert practitioners have developed through years of human-automation interaction experience must be formally taught to cadets who enter the profession in an AI-augmented environment where that experiential development pathway will be compressed. The AI4People framework's [4] autonomy principle — AI should preserve rather than undermine human decision-making capacity — is operationalized in practice by experienced practitioners through exactly the conditional trust architecture this study documents; the educational challenge is to develop this conditional trust architecture through curriculum before cadets encounter the professional contexts in which its absence creates ethical and safety risk.

The Accountability Asymmetry theme's finding that professional liability uncertainty is a primary driver of practitioner caution about AI recommendation adoption — more consequential than performance skepticism — reveals a dimension of AI adoption resistance that the conventional TAM-based [7] adoption framework does not capture. TAM predicts that adoption is determined by Perceived Usefulness and Perceived Ease of Use; the Accountability Asymmetry finding suggests that for professional maritime contexts where legal liability is a live professional concern, Perceived Accountability Clarity may be an additional and independent adoption determinant not anticipated by standard technology acceptance theory. This finding extends the human-AI ethics literature's theoretical account of adoption determinants in high-consequence professional contexts and suggests that regulatory clarity on AI accountability frameworks, rather than merely technical performance improvement, may be the most productive intervention for increasing appropriate maritime AI adoption.

The Generational Ethics Gradient theme's spontaneous cross-program emergence — offered by experts without FGD protocol prompting — suggests that practitioner concern about the next generation's human-AI ethical calibration is a salient, actively discussed professional community concern rather than an incidental observation. This concern is empirically warranted by the companion DIGITS Paper 1 finding: Cadet AI adoption Behavioral Intention scores (overall M=3.54) were higher than their AI Awareness scores (M=3.57 overall but lower for DN), suggesting that cadets are enthusiastic about AI adoption at levels that their actual AI knowledge does not fully support — a pattern consistent with the lower autonomy reflex the Generational Ethics Gradient theme identifies. STIP Jakarta is institutionally positioned to address this gradient by developing explicit human-AI ethics instruction within its AI literacy curriculum — not to suppress AI adoption enthusiasm but to develop the conditional trust architecture that makes adoption professionally appropriate rather than naively uncritical [2].

4. Conclusion

This study has documented four thematic categories structuring STIP Jakarta expert practitioners' ethical orientation toward human-AI interaction in maritime decision-making: Conditional Trust Architecture (28/30), Human Override Norms (27/30), Accountability Asymmetry (25/30), and Generational Ethics Gradient (22/30). DN experts exhibit the most human-primacy oriented ethical positioning across all themes, consistent with the navigation professional formation context's emphasis on Master's authority and COLREGS compliance as non-delegatable professional responsibilities. The Accountability Asymmetry theme's identification of professional liability uncertainty as a primary adoption constraint extends TAM-based adoption theory in an important direction for high-consequence professional contexts. The Generational Ethics Gradient's spontaneous cross-program emergence identifies the next generation's AI trust calibration development as the most institutionally urgent and currently unaddressed human-AI ethics challenge STIP Jakarta faces. These findings contribute the first practitioner-grounded, ethics-framework-informed account

of human-AI interaction values in Indonesian maritime professional contexts and provide direct curriculum development priorities for responsible AI integration in maritime vocational education.

5. References

[1] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, "An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," Minds Mach., vol. 28, no. 4, pp. 689–707, 2018. DOI: 10.1007/s11023-018-9482-5

[2] K. Bogusławski, M. Gil, J. Nasur, and K. Wróbel, "Implications of autonomous shipping for maritime education and training: the cadet's perspective," Marit. Econ. Logist., vol. 24, no. 2, pp. 327–343, 2022. DOI: 10.1057/s41278-022-00217-x

[3] International Maritime Organization, "Maritime Autonomous Surface Ships (MASS) — Outcome of the regulatory scoping exercise," MSC-MEPC.3/Circ.7, IMO Publishing, London, 2021.

[4] L. Floridi et al., "An ethical framework for a good AI society," Minds Mach., vol. 28, no. 4, pp. 689–707, 2018. DOI: 10.1007/s11023-018-9482-5

[5] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," IEEE Trans. Syst. Man Cybern. — Part A, vol. 30, no. 3, pp. 286–297, 2000. DOI: 10.1109/3468.844354

[6] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nat. Mach. Intell., vol. 1, no. 9, pp. 389–399, 2019. DOI: 10.1038/s42256-019-0088-2

[7] J. W. Creswell and C. N. Poth, Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th ed. Thousand Oaks, CA: Sage Publications, 2018.

[8] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qual. Res. Psychol., vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa

[9] Y. S. Lincoln and E. G. Guba, Naturalistic Inquiry. Beverly Hills, CA: Sage Publications, 1985.

[10] A. Wahid, M. Y. Jinca, T. Rachman, and J. Malisan, "Determination of Indicators of Implementation of Sea Transportation Safety Management System for Traditional Shipping Based on Delphi Approach," Sustainability, vol. 15, no. 13, 2023. DOI: 10.3390/su151310080

[11] F. Mutohhari, P. Sudira, F. D. Isnantyo, and N. W. A. Majid, "Generic green skills: maturity level of vocational education teachers and students in Indonesia," Int. J. Eval. Res. Educ., vol. 14, no. 1, pp. 179–187, 2025. DOI: 10.11591/ijere.v14i1.29191

[12] L. Widaningsih, S. Rahayu, V. Dwiyanti, and W. Widanengsih, "Analysis and Mapping of Technical Competency Needs for TVET Teachers in the Era of Industry 4.0," J. Tech. Educ. Train., vol. 17, no. 2, pp. 151–164, 2025. DOI: 10.30880/jtet.2025.17.02.011

[13] H. Li, S. I. Khattak, X. Lu, and A. Khan, "Greening the Way Forward: A Qualitative Assessment of Green Technology Integration and Prospects in a Chinese Technical and Vocational Institute," Sustainability, vol. 15, no. 6, 2023. DOI: 10.3390/su15065187

[14] R. S. Wulandari and T. Cahyadi, "Engineering Technology and Education in Maritime Studies: A Qualitative Examination," J. Eng. Sci. Technol., vol. 20, no. 2, pp. 28–35, 2025.

[15] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, 1989. DOI: 10.2307/249008