HARVARD | BUSINESS | SCHOOL

Golden Opportunity? Voluntary Sustainability Standards for Artisanal Mining and the United Nations Sustainable Development Goals

Kristin L. Sippl

Working Paper 19-024

Golden Opportunity? Voluntary Sustainability Standards for Artisanal Mining and the United Nations Sustainable Development Goals

Kristin L. Sippl
Harvard Business School

Working Paper 19-024

Golden Opportunity?
Voluntary Sustainability Standards for Artisanal Mining and the United Nations Sustainable Development Goals

7,998 words

Despite decades of effort, the international community still struggles to balance economic, social, and environmental objectives. In 2000, the United Nations launched eight Millennium Development Goals to guide global efforts, but made only mixed progress towards achieving them (Kanie and Biermann 2017). In 2015, the United Nations launched 17 Sustainable Development Goals (SDGs) to update the global agenda by addressing ongoing issues such as poverty, human and environmental health, education, gender equality, infrastructure, peace and the need for global partnership (UN 2019).

Many of the actors working to contribute to SDG attainment are transnational, meaning they derive authority from civil society or the private sector and operate across national borders. Third-party product certification and labeling programs are one type of transnational actor explicitly claiming to contribute to goal attainment (de Pena 2015, FSC 2019, MSC 2019). These programs require value chain actors to obtain independent audits certifying compliance with programs' voluntary sustainability standards, then connect the resulting product to political consumers willing to pay more for sustainable development. Stronger SDG contributions come from programs with more goals shared with the SDGs, more stringent standards, wider spread program uptake by targeted actors, and more positive indirect effects. Such indicators alone do not constitute contributions due to the array of implementation problems plaguing all governance programs (Selin et al. 2018, Short et al. 2016). But they do serve as necessary conditions for program contributions, and are therefore important to study (Auld et al. 2008, Gulbrandsen 2009, Kalfagianni and Pattberg 2013, van der Ven and Cashore 2018).

The certification literature is accruing deep knowledge of programs in a few critical sectors, such as forestry, fisheries, and tropical agriculture (Auld 2014). Research is only beginning to broaden analysis to sectors in which certification is missing (e.g. plastics, pets) or nascent (e.g. cannabis, minerals) despite the serious sustainability challenges these sectors present (Auld et al. 2018, Bennett 2018, Bloomfield 2015, Bloomfield and Schleifer 2017). For example, artisanal and small-scale gold mining (ASGM) is one of newest sectors targeted by certification programs. As the United Nations Environment Program (2018) reports, ASGM employs minimally 15 million people in developing countries who often work informally, employ children, degrade forests and waterways, and use toxic chemicals that harm humans and the environment both locally and globally (e.g. ASGM is a leading cause of global mercury pollution). These are typically actions of last resort—miners are driven to ASGM by poverty in contexts where alternative livelihoods are scarce. The international community recognizes the importance of addressing ASGM and its link to the SDGs (Selin 2014, IISD 2019). Yet research on certification program performance in the sector is currently limited. This paper therefore asks: What is the strength of ASGM certification program contributions to the SDGs? What are the pathways of program influence, and magnitude of program effects?

Answering these questions contributes to research on transnationalism in several ways. First, it connects certification to the SDGs, a linkage common in policy spheres but largely missing from the academic literature and important due to the different abilities certification programs may have relative to other governance actors (Bennett 2018a). Second, the paper makes this contribution through analysis of ASGM certification, an understudied sector-case that provides insight into the challenges certification faces as it diffuses to non-traditional sectors (Childs 2014, Fisher 2018, Hilson 2008, Hilson et al. 2016). Finally, this is the first academic analysis to provide a comprehensive comparison of current program contributions. Taken together, the findings contribute to debates on global governance fragmentation and its impact on regime effectiveness (Biermann et al. 2009, Young 2011, Zelli and Van Asselt 2013).

The paper begins by situating the ASGM case in debates on global governance fragmentation and certification program effectiveness. In doing so, it derives a framework to analyze program contributions to the SDGs. It then presents the findings underpinning the paper's argument: ASGM certification

programs are making only weak to moderate contributions to the SDGs because the positive effects of high SDG alignment and moderately stringent rules are offset by weak uptake patterns and indirect effects. The paper concludes by highlighting ways programs could improve and associated paths for future research.

Transnationalism and the SDGs

Global governance regimes are collections of actors working to achieve shared goals. As transnationalism increases, regimes become more fragmented, with debatable impacts on their effectiveness (Biermann et al. 2009; Zelli and Asselt 2013). Young (2011) summarizes the optimistic view that hard law is not necessary for positive regime effects. Regime design should adapt to fit the targeted problem, and because transnational actors are more flexible than other actors types, their presence makes regimes more agile. Further, transnational actors can offset deficiencies in public actors, experiment with new governance approaches, inspire race-to-the-top dynamics, serve as links between regimes, and become stronger through regime participation, launching virtuous governance circles (Cashore et al. 2007, Green and Auld 2017, Hoffmann 2011).

Pessimists counter that fragmentation causes coordination problems and turf tensions within regimes (Young 2011). Race-to-the-*bottom* dynamics may prevail, non-profit organizations sacrifice best practices to win donor contracts, and transaction costs increase due to forum shopping (Cooley and Ron 2002, Schleifer et al. 2018). Other concerns are more fundamental. When fragmentation is driven by businesses, trade associations, or market-driven civil society actors, legitimacy and power deficiencies can be amplified, and the market is relied upon to solve the very problems it created (Bernstein 2004, Zelli and Asselt 2013).

Certification Programs and the SDGs

Certification programs are transnational actors explicitly claiming membership in the SDG regime. Fairtrade International claims to contribute to 8 SDGs and that SDG1 (reducing poverty) is "central to Fairtrade's mission" (de Pena 2015). The Forest Stewardship Council claims to contribute to 11 SDGs, while the Marine Stewardship Council claims alignment with 8 (FSC 2019, MSC 2019). Analyzing an aggregate of 241 voluntary programs, the United Nations Forum on Sustainability Standards found that programs are "best positioned to directly contribute" to 10 SDGs, and particularly to SDG12 (responsible commerce), SDG8 (economic growth), and SDG15 (life on land) (UNFSS 2019).

These studies have three deficiencies. First, self-evaluations lack objectivity, and the methodologies programs used are opaque. Second, the UNFSS's relatively objective assessment does not analyze the full content of individual standards, which biases results and limits their usefulness. Third, these analyses look only at the program's content and not its uptake levels which, as the next section explains, are critically important to estimating program effects. This paper helps address these deficiencies by objectively and thoroughly assessing programs in a way that is replicable by future researchers.

Evaluating Certification Program Effects

Measurements of certification program effects are sensitive to the definitions used and problems studied. Young (2011) notes that positive effects exist on a continuum from solving the problem motiving action to generating useful information. To measure certification program effects, social science research typically foregoes biophysical testing, focusing instead on indicative conditions for effects. As demonstrated by Auld et al. 2008, Gulbrandsen 2009, and Kalfagianni and Pattberg 2012, most studies begin by assessing the comprehensiveness and stringency of a program's rules, i.e. the breadth of sustainability issues addressed and the degree to which rule compliance solves the problem. Next, most studies measure program uptake patterns to assess whether the targeted actors are participating in

programs. All studies note the inverse relationship between rule stringency and uptake, and the resulting difficulty of creating meaningful change. The 'selection problem' means that more stringent rules attract targets already performing well, and less stringent rules attract worse-performing targets but require less change. Studies begin to depart in the weight they give to unintended and spillover effects, i.e. surprise impacts of programs on participants and the environment, and impacts on non-participating communities and goals. Few go as far as Kalfagianni and Pattberg (2012), who add measurements of problem structure, access to decision-making procedures, and quality of audits to their analysis.

This literature suggests that programs have mixed effects across sectors. In some sectors (e.g. cannabis, aquarium fish) certification has yet to emerge or failed quickly after its launch (Bennett 2018, Bloomfield and Schleifer 2017). In forestry and fisheries, certification has positive effects in industrialized countries, but neutral to negative effects in developing countries, where uptake among low-income producers is low, and tropical forest conversion and biodiversity loss continue (van der Ven and Cashore 2018). Effects in the coffee sector vary by country. Certification drove higher wages and investments in public goods in Guatemala yet the same or worse wages in Mexico (Jaffe 2014, Linton 2015). In Uganda, programs tend to achieve either economic or environmental goals, but not both (Vanderhagen et al. 2018). Looking across agricultural sectors, Schleifer et al. (2018) argue programs are not very "producer-friendly." And looking across manufacturing sectors, Locke 2013 concludes that programs alone are insufficient to solve problems, but that they do bring benefits through training and capacity building. Despite most studies focusing on these sectors, research on new sectors is growing (van der Ven et al. 2018). This paper joins these efforts by examining the effects of certification programs in the ASGM sector.

Gold Jewelry Certification

As Author's (2015) overview explains, gold jewelry sourced from certified ASGM mining operations first hit global markets in 2011. It was labeled "Fairtrade and Fairmined Gold," signaling its derivation from a partnership between Fairtrade International (FLO) and the Alliance for Responsible Mining (ARM). ARM formed in 2004 to unite activists and artisanal miners wanting to make and sell more environmentally-friendly gold. In 2009, ARM signed a 3-year contract with FLO to further these aims, retrofitting FLO's agricultural standards to the ASGM context. Miners complying with the program's social and environmental rules received 95% (rather than the typical 70%) of the international gold price plus a 10%-15% premium to be invested in their communities.

By 2013, program feedback had accrued: miners wanted more money, consumers wanted lower prices, businesses wanted higher volumes. ARM believed raising premiums and mass-balancing (mixing certified and uncertified gold for higher volumes and lower prices) was the best response. FLO disagreed, citing mass-balancing as 'greenwashing' and wanting to increase miners' incomes by *lowering* premiums to attract more customers. Unable to resolve these differences, the organizations split to create the competing programs that exist today: "Fairtrade Gold" (FLO) and "Fairmined Gold" (ARM). Besides changes on mass-balancing and premiums, the revised programs (created in consultation with at least one miner) have much in common. Both aim to create "opportunities" for artisanal miners in low- and middle-income countries, cover both social and environmental issues, increase rule stringency over time, and target the same consumers (FLO 2013, ARM 2014). Programs differ in their views on capitalism. FLO seeks to "change conventional trading systems" (FLO 2013, p.4) whereas ARM aims to build ethically "viable businesses" (ARM 2014, p.4).

The literature on ASGM certification primarily focuses on the original joint-program. Hilson et al. (2016) worried that the joint program would not reach the miners most in need of help. If they *were* reached, Childs (2014) worried that miners might reject certification due to their negative experiences with past development interventions and reliance on informal networks for services. And if miners *wanted* to certify, Fisher (2017) charts the obstacles they must overcome to pass audits, while Author (2015) and Hilson et al. (2016) warn that governments might prefer to give permits to larger, more lucrative mining operations. Hilson (2008) throws doubt on the entire premise of certified ASGM, noting

that in many regions the final buyer is governments, not jewelry consumers. While helpful, this literature needs the update that this paper provides via comparative analysis of current program performance globally with recommendations for program improvement.

Data and Analysis

Case selection for the paper was straightforward. As of mid-2019, the International Trade Center's Voluntary Sustainability Standards database identified only two certification programs targeting ASGM: ARM and FLO. They are therefore the paper's sample. To assess their contributions to the SDGs, this paper follows the literature by measuring four indictors of effects: SDG alignment, rule stringency, uptake patterns, and indirect effects.

SDG alignment is how well certification rules are positioned to contribute to the SDGs. Following the UNFSS's (2018) protocol, the author assessed each certification rule and matched its likely effects to a corresponding SDG (if matches existed). For example, a rule requiring mercury-reducing technology matches SDG3 (human health), but its effects on SDG4 (education) are indirect and SDG9 (infrastructure) nonexistent, so the rule is not coded as matching the latter two. The number of SDGs with at least one matching rule was counted. The more matches, the higher program alignment. SDG alignment proxies for measures of rule comprehensiveness in the literature, and results are believed to be easily replicable.

Rule stringency is the degree to which the behavior change required by a rule solves a problem, and is coded as high, moderate, low, or none. With regard to mercury, for example, a highly stringent rule requires mercury elimination, a moderate rule requires reduction, a low-level rule requires aspiration to reduce, and if a standard is missing rules on mercury this is coded as 'none.' Four types of uptake patterns are measured. Producer type measures the poverty status of participating miners upon program entry using the World Bank's (2018) international poverty line of $1.90 per day (extreme poverty) as well as higher lines at $3.20 and $5.50 per day signifying moderate poverty and poverty-free statuses. Adoption type is whether more or less stringent versions of standards are adopted (if options exist). Adoption levels are the numbers and percentage of targets participating in programs. Adoption trends are changes in adoption levels over time. Finally, the analysis measures indirect pathways through which programs might contribute to goals (called indirect effects for brevity). This indicator captures the array of unintended and spillover effects measured in other studies, such as effects on targets who choose not to participate, government regulations, and certification programs in other sectors (Auld et al. 2008).

Indicators are measured using data from several sources. First, program websites publish their standards, adoption data, and premium expenditures. The author collected adoption data 2013-2019 (6 years). The resulting database of adoption trends (i.e. certifications and decertifications) is unique and important because certification status is only available for the current month, making historical data unavailable to the public. The database's accuracy was confirmed via emails with the programs, which also confirmed the number of certified miners and amounts of certified gold sold. Interviews with program representatives and academics supplement the research, as does the author's observations of certification program behavior at United Nations mercury treaty conferences in 2012 and 2018.

ASGM programs making the strongest contributions to the SDGs would have high SDG alignment, highly stringent rules with widespread and increasing adoption by the poorest and most environmentally harmful miners, widespread and increasing adoption by jewelry retailers, and helpful indirect effects. Audit quality is not included in the analysis because data was not available, so results reflect a best-case scenario of perfect compliance. Accordingly, if miners were mid-audit at time of writing, they were counted as certified. And problem structure and influence over rule-making are addressed in the conclusion to keep the analysis concise.

Results

Overall, ASGM certification programs are making weak to moderate contributions the SDGs, but they are well positioned to make stronger ones. This section presents the findings on the four indicators that suggest this result. Final evaluations are summarized in Table 1.

Table 1. Evaluations of SDG Contribution Strength				
	SDG Alignment	**Rule Stringency**	**Uptake Patterns**	**Indirect Effects**
ARM	High	Moderate	Weak	Weak
FLO	High	Moderate	Weaker	None

SDG Alignment

Expecting programs to align with all 17 SDGs is unrealistic. Nevertheless, the more alignments, the better for contribution strength. Rule-to-goal matching revealed that programs are aligned with 10 SDGs (Table 2). This is higher than the alignment numbers in ARM's and FLO's self-assessments (9 and 8 respectively) and the mix of goals is different. The paper and programs agree that programs align with goals on poverty, gender, economic growth, and responsible commerce. The paper disagrees with program claims that they align with goals on hunger, climate change, and partnerships. Although FLO's agricultural programs do align with the hunger goal, there is nothing in either program's mining rules that addresses hunger except indirectly through poverty reduction. The only connection to climate change is through forest restoration, but mining requires continuous acquisition of (often forested) land—on balance rules encourage more forest loss than gain. And partnership requirements are missing from programs rules except in the case of miners applying for exceptions to rules (which is too obscure to count as a direct match). For other goals, the paper gives more credit than the programs give themselves. The paper argues that both programs' mercury rules align with goals addressing the health of humans (FLO does not) and life below water (neither program does). The paper further argues that programs' conflict rules align with the peace goal (ARM does not). In addition to presenting this paper's views on alignments, Table 2 highlights the SDGs for which program rules are particularly stringent (and not), with the rule stringency analysis presented in the next section.

Table 2: Program Alignment and Rule Stringency Relative to the SDGs				
			Alliance for Responsible Mining (ARM)	**Fairtrade International (FLO)**
Sustainable Development Goals		**Certification Program Goals**	**Program Requirements**	
Responsible Commerce*	Economic Growth*	Legalization	Acquire legal permits	Acquire legal permits
		Child Labor	Bans children 7-14, but fewer programs	Allows children 7-14, but more programs
	Gender Equality*	Gender Equality	Non-discrimination in hiring	Non-discrimination in hiring
	Peace	Reduce Conflict	Bans fewer conflict types	Bans more conflict types
	Life on Land	Ecological Restoration	Re-vegetate	Re-vegetate
		Critical Ecosystems	No impact assessment, no third parties required; focus on livelihoods; nothing on poaching	Requires impact assessment and third-party approval; focus on environmental effects; nothing on poaching

	Clean Water	Water Management	No dumping, end acid drainage; turbidity caps	No dumping, end acid drainage; turbidity caps
	Health Life Below Water	Toxic Substances	Retorts (Basic) or mercury-free technology (Ecological); amalgam cyanide leaching banned Year 3	Retorts (Basic) or mercury-free technology (Ecological); amalgam leaching banned Year 0
	No Poverty˜	Development: Producer Uptake	Cash: 95% gold price Account: $4,000 (Basic) Account: $6,000 (Ecological) Cost-sharing: None	Cash: 95% gold price Account: $2,000 (Basic) Account: 15% (Ecological) Cost-sharing: None
	Reduced Inequality*	Development: Buyer Uptake	Fully-traceable: Fairmined Labeled, license fee	Fully traceable: Fairtrade Labeled, license fee
			Semi-traceable: Fairmined Incorporated (smaller fee)	Semi-traceable: Fairtrade Sourcing (smaller fee), Registered Goldsmiths (no fee)
			Not traceable: Fairmined Certificates, non-market-based donations	
*High stringency; ˜ Low stringency				

Rule Stringency

Program rules targeting the 10 SDGs in Table 2 vary in both their absolute and relative stringency. Programs begin by stipulating who can mine. When forming ASMOs, miners cannot discriminate against women (highly stringent on SDG5) and must acquire legal permits, i.e. exit the informal economy and contribute to formal economic growth (highly stringent on SDG8, especially compared to programs or laws encouraging ASGM bans). Children frequently mine alongside their families to help generate income, despite the risks to their health, safety, and education this poses (UNEP 2018). Programs allow children aged 0-7 to work with their families due to the lack of childcare in communities. Programs differ with regard to children age 7-14: ARM bans them from mining while FLO allows them to mine if ASMOs also create child services programs. These differences reflect debates about best methods for combatting child labor. Bans protect children from threats posed by mining, but may expose them to other threats and reduce the ability of families to pay for things such as school fees. Allowing children to remain on mining sites endangers them, but the extra income may help families invest in safer mining methods or transition to safer livelihoods. Therefore both rules are deemed highly stringent on the aspects of SDG8 (economic growth) that address child labor, and their relative stringency is deemed subjective because it depends on beliefs about strategies.

Next, programs stipulate *where* mining can take place. In alignment with SDG16 (peace), both programs ban mining in areas of civil conflict or conflict with agriculturalists. FLO is more stringent than ARM on peace because FLO additionally bans mining near indigenous communities and large-scale mining operations. Yet both programs allow miners to apply for exceptions enabling them to mine in such areas, weakening rule stringency from high to moderate. Both programs further ban mining in critical ecosystems and protected areas (SDG15), but again offer exceptions that weaken rule stringency to moderate at best. From a relative perspective, FLO is more stringent than ARM on protected areas because it requires ASMOs to conduct environmental impact assessments and gain third-party verification that their mining will be environmentally benign. ARM requires neither. ASMOs instead must prove that there are no alternative livelihoods in their vicinity. While these exceptions hurt SDG16 and SDG15, more mining means more income, which helps SDG10 (reducing inequality) and SDG8 (economic growth).

In addition to specifying who can mine and where, programs stipulate *how* mining can be done. Both programs offer two types of standards: a more lenient basic standard and a stricter "Ecological"

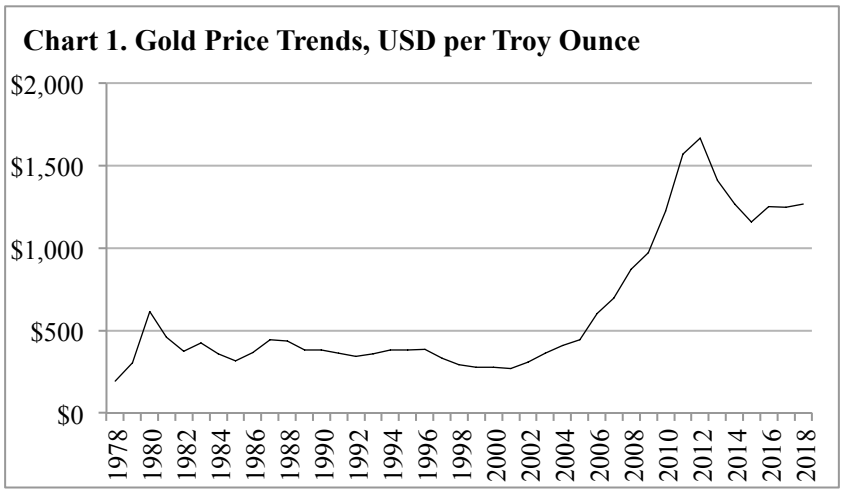
standard. Continuing with rules aligned with SDG15 (life on land), programs require miners to re-vegetate mined areas after three years (ARM) or six years (FLO) in the program (both require it earlier in Ecological standards). But re-vegetation does not fully restore deforested areas, and mining often necessitates deforestation because of where gold is found and the need to constantly acquire new gold-laden land. Completely absent from programs are rules protecting wildlife, which is harmed by forest degradation, toxic chemicals, and miners hunting animals for food. Taken together with the exceptions enabling mining in protected areas, rules are deemed to have moderate to low stringency on SDG15 (life on land), reflecting the inherent tensions between sustainability goals and non-renewable resources like gold.

Rules on SDG6 (clean water) similarly lack stringency. As an entry requirement miners must stop dumping fuel and contaminated water into or near waterways, which is a strong rule. But miners have three years to fix the problem of acid mine drainage, a process in which exposed rocks create toxic chemical compounds that seep into drinking water. This is a long time for pollution to continue. Both programs lack rules on sanitation in mining sites, which are often remote and rudimentary. ARM does cap water turbidity, making their rules more stringent than FLO's. But at an absolute level, these omissions and delays make rules on SDG6 (clean water) moderately stringent, at best.

Programs do slightly more for SDG3 (health) and SDG14 (life below water), mostly by addressing mercury[1] (neither program addresses malaria and its link to the pools of water mining creates). Programs are deemed moderately stringent because they require miners to adopt technologies that reduce mercury (basic standard) or eliminate it (Ecological standard) (an assessment of high stringency is reserved for programs focused solely on elimination). The retorts required by the basic standard reduce mercury by 75-95% if used correctly, but they seldom are, which can make pollution worse by concentrating vapor (AGC 2013). They also keep demand for mercury alive when the international community wants to reduce or eliminate it. In addition to technology, FLO requires miners to not mix cyanide with mercury as an entry requirement, whereas ARM allows this until miners' third year in the program, making its rule less stringent because cyanide heightens mercury's harm.

A final way programs effect mercury is through the incentives they offer for certifying under the basic versus ecological standard. Both programs pay miners the same wage—95% of the international gold price—but different in their social premium payments. ARM pays twice as much as FLO for certification under the basic standard: $4,000/kg compared to $2,000/kg. The payments for Ecological standards vary according to the gold price: ARM pays $6,000, but FLO pays 15% the international gold price. Programs pay the same amount for Ecological certification when the gold price is $40,000/kg ($1,244/troy ounce). When the price is above this, FLO pays more; when it is below, ARM pays more. As Chart 1 shows, FLO's incentives for Ecological gold (and therefore mercury elimination) were higher in all years except 2015. Based on this, it is tempting to conclude that FLO is more stringent on mercury. But it is more accurate to say that the effects of incentives depend on the gold price and on beliefs about incentivizing lower-mercury versus mercury-free gold. ARM, for example, might be aiming for mercury reductions equivalent to FLO's but through many small reductions (basic standard) rather a few big ones (Ecological standard).

[1] Mercury bio-accumulates in aquatic ecosystems (UNEP 2018).



Data: World Gold Council, 2019

Certification payments also align with SDG1 (zero poverty), SDG10 (reduced inequality) and SDG8 (economic growth). In terms of stringency, the 36% wage boost from certification is a highly stringent rule spurring economic growth and reducing inequality, but a low-stringency rule on poverty reduction. The average miner produces 0.02kg of gold per day for an income of $1.89 per day (extreme poverty). If the average miner certified, their wages would increase from $1.89 to $2.57 per day: helpful, but not enough to lift them above the moderate poverty line ($3.20 per day) or poverty-free line ($5 per day). Therefore program rules do not eliminate poverty, even if social premiums (which are not wages) are factored in. If social premiums are considered, ARM is more stringent than FLO on spurring growth and reducing inequality because their basic standard pays twice as much, but amounts are still not high enough to lift the average miner out of poverty.

These income-related SDGs are further served by incentives targeting the buyer-side of the value chain, i.e. the three trading models created to convince manufacturers, wholesalers, and retailers to participate in programs. In addition to fully traceable and labeled gold, both programs have semi-traceable mass-balancing models that track gold from the mine to the refinery, but no further. Gold sourced through these models is unlabeled, less expensive and available in larger volumes. FLO offers a second semi-traceable model identical to their mass-balancing model except it aims to attract small wholesalers and retailers by waiving the prior model's licensing fees. ARM likewise differentiates itself via its third model, which aims to attract donors who would otherwise forego program participation for reasons such as branding or industry position. This model completely separates the donation from gold sourcing. FLO does not have an equivalent model, believing it departs too far from their 'trade, not aid' philosophy and allows harmful industry sourcing practices to continue (Author's interview with FLO, 2014). Such models do, however, send income to miners that might otherwise be missed. As with rules on child labor and mercury, the relative stringency of rules on trading models is subjective, based on beliefs about best strategies. But since programs clearly have models to attract every type of business, they are deemed to support the rules' assessment as highly stringent on inequality and growth.

When all program rules are evaluated together, they are moderate to highly stringent on SDG12 (responsible commerce). As Table 2 showed, they are highly stringent on SDGs related to economic growth, gender equality, and reducing inequality; they are moderately stringent on clean water, peace, and life on land. They have moderate to low stringency on health and life below water, and low stringency on poverty. Table 3 shows rules' relative stringency to the degree it can be assessed. The remaining factors shaping SDG contributions—uptake patterns and indirect effects—are discussed next.

Table 3. Relative Rule Stringency	
Equal Stringency	Responsible Commerce, Gender Equality, Economic Growth (Child Labor)
ARM is More Stringent	Poverty, Inequality, Economic Growth, Clean Water
FLO is More Stringent	Peace, Health, Life on Land, Life Under Water

Uptake Patterns

Adoption levels, types, and trends are key components of the analysis. In programs' 10th year of working with miners, 19 ASMOs (1,900 miners) are certified, equivalent to 0.01% of the ASGM population. ARM and FLO have 10 ASMOs certified each (1 ASMO is certified under both programs). ARM has 7 certified under its basic standard and 3 under its Ecological standard. FLO has 10 certified under its basic standard and none under its Ecological standard. When comparing programs, it is important to note two things. First, it is possible that programs have the same number of ASMOs but different numbers of miners (FLO lacked data on the number of miners per ASMO, so estimates are based on ARM's averages). Second, adoption trends fluctuate significantly over time.

Table 4: Current Adoption Levels and Types (Spring 2019)					
ASMOs Selling Certified Gold			**Licensees Selling Certified Gold**		
	ARM	**FLO**		**ARM**	**FLO**
Bolivia	1	0	Africa	0	6
Colombia	3	0	Asia	6	1
Kenya	0	1	Australia	1	4
Mongolia	2 (2 Ecological)	0	Europe	119	46
Peru	4 (1 Ecological)	9	North America	53	3
			South America	17	0
Global Totals:	10 (3 Ecological)	10 (0 Ecological)		196	60

When the joint-program was active (2011-2013), 4 ASMOs participated (3 basic, 1 Ecological) and none of these ASMOs are still certified today. When miners had a choice between revised programs (2013-2019), 19 ASMOs gained and maintained certification with at least one program, while 7 ASMOs tried certification but eventually decertified (a 27% drop-out rate). ARM gained 10 and lost 4, while FLO gained 10 and lost 5. Narrowing analysis to Ecological ASMOs, ARM gained 3 and lost 1, while FLO did not gain any. It is further important to note that the rates of gains and losses are not steady over time. For example, as recently as early 2018, there were a total of 10 ASMOs certified (roughly half the number certified today): ARM had more than twice as many as FLO (8 to 3), and ARM's dropout rate was 30% compared to FLO's 57% (FLO had lost more ASMOs than it had gained). If FLO's new ASMOs maintain certification, then their gold program growth is equivalent to FLO's certified wine growth (adding, on average, 10 producer groups every five years) yet behind coffee (which adds, on average, 75 producer groups every five years) (FLOCERT 2017).

The final uptake pattern studied is participant type. Both programs aim to reduce poverty, but the miners participating in programs were, on average, living *above* poverty lines prior to certification. While the average artisanal miner sells about 0.02 kilograms of gold per day resulting in $1.89 per day (extreme poverty), the miners participating in programs mine ten times more, about 0.20 kilograms per day resulting in $14.84 per day (roughly three times above the $5 per day poverty line in middle-income countries). This finding confirms worries in the literature—programs are not reaching their intended targets. It further casts doubt on assumptions that mercury use is purely poverty-driven.

Producer uptake findings are critical indicators of program effects, but producer uptake will not last without ongoing support from retailers. 2013-2017, response from retailers was slow but positive—ARM and FLO gained equal numbers of stores (136 and 137, respectively) with relatively few store dropouts. 2017-2019, however, ARM's numbers grew while FLO's plummeted. Led by gains in Europe

and North America, ARM's total climbed to 196 stores. Led by losses in Europe, FLO's total fell to 60 stores (slightly below their 2016 levels).

Taken together, these uptake patterns inspire cautious optimism. Optimism is appropriate for four reasons. First, programs accurately detected that the joint-program was unpopular and made revisions. Second, the 29% dropout rate 2013-2019 is lower than the estimated average of 40% in the coffee sector (Interview with academic Janina Grabs, 2018). Third, the number of Ecological certifications rose in recent years. And fourth, FLO's recent surge in ASMO numbers might reflect program learning.

Caution, however, should temper optimism. It is not clear whether FLO's recent surge is exponential growth or temporary volatility. If FLO loses its recent gains, their growth will clearly be lagging behind both ARM and other certified sectors. Also worrisome is the low levels of miners certified to date: change in 0.01% of the target population is not enough adoption to make a strong contributions to the SDGs. Worse yet is the low number of miners adopting Ecological programs, despite their not being in poverty prior to or during participation. One might hope that the predominance of basic standards simply reflects that time is needed to save for mercury-free investments, i.e. development will lead to environmental protection. Unfortunately, despite the $1.2 million (FLO) and $1.7 million (ARM) generated from programs' social premiums 2013-2017, only one ASMO made mercury-free investments: Colombian cooperative Iquira invested 60% of its premiums in mercury-eliminating processing plants (ARM 2019). ASMOs in Peru invested in water treatment plants, educational facilities, a vehicle, and two sports fields. Colombian miners invested in safety equipment, micro-loans, and a computer. Such investments enhance morale, capacity and productivity, but also show that higher incomes do not necessarily lead to mercury reductions.

Indirect Effects

The final indicator of SDG contributions is indirect effects. Fisher (2017) confirms that ideas about best practices are indeed spreading from certified to uncertified mining organizations, helping to build capacity among miners who have yet to certify. But both she and the author's interviews with FLO in 2015 also confirm the concerns raised by Childs (2014): miners are slow to trust certification program representatives due to negative experiences with such interveners in the past, and are becoming disheartened by working towards yet failing to achieve certification (the case in several lower-capacity mining communities in Africa).

Indirect effects of private governance can also flow through effects on public governance. One pathway is through impacts on government representatives responsible for ASGM policy. Sun (2017) reports the impacts of transnational actors on delegates to and thereby the text of the UN Minamata Convention on Mercury. Yet certification programs are not mentioned in his analysis, and this paper's author likewise observed the relatively muted role in international meetings and conferences programs play relative to other actors. In 2012, both ARM and FLO were present but reserved at a major Minamata Convention negotiating meeting. At the most recent conference of the parties in 2018, ARM (but not FLO) was present and played more active role than in the past (e.g. presenting their program to delegates in an official conference side event). Still, relative to the extremely active approach taken by capacity-building organization the Artisanal Gold Council (e.g. bringing delegates and program representatives together during breaks and parties to discuss current and future projects) there is no evidence (yet) that certification programs are effecting public policies through government contact.

A second pathway of influence on public policy could be through political consumers. Raising awareness among voters about sustainability issues through programs like ASGM certification could build stronger constituencies for stronger regulations. Yet the evidence for this is also weak. In a 2015 survey of FLO customers, only 3% were aware that certified gold existed compared to the roughly 50% who were aware of certified chocolate and 64% aware of certified coffee and tea (Ingle and Rhodes 2016). Of course, coffee certification has existed for nearly 30 years compared to gold's 8 years on the market. Still, at this time it does not appear that gold certification is contributing to the SDGs by raising a constituency for stronger gold regulations.

Conclusion

To summarize, ASGM certification programs are making weak to moderate contributions to the SDGs. The positive effects of high SDG alignment and moderately stringent rules are offset by weak uptake patterns and indirect effects, casting doubt upon the magnitude of fragmentation's benefits. ARM's and FLO's strategies for contributions differ. ARM strives for widespread uptake via its high-paying basic standard, trusting that consumers will pay the high prices and development will cause environmental upgrading among miners. FLO believes customers are more price sensitive, so keeps payments for basic standards low and payments for Ecological standards high to maximize incentives for upgrading. These strategies have resulted in low levels of uptake and only among non-impoverished miners who mostly choose the weakest standards and do not invest in mercury-free technology. Indirect program effects are minimal, although along with the other indicators they highlight important problems to solve. Programs now face a crossroads: they should either improve or disband.

If they wish to improve, three changes and associated paths of academic research might help. First, ARM and FLO could reunite to form a single certification program. Consumers already face an overwhelming number labels, and educating them about the need for certified gold will be hard enough without adding competing programs to the mix. Further, the mass-balancing issue that once divided programs is resolved (they now both agree it is necessary). Finally, as fragmentation critics might predict, program competition is positioned to inadvertently harm mercury reduction efforts. If programs compete in the same mining community, ARM's basic standard (offering medium-reward for low-effort) will likely lure some medium-capacity miners away from FLO's ecological standard (offering only slightly higher reward for *much* higher effort). Yet under a FLO monopoly, FLO's basic standard pays so little that many miners may find the extra effort worth it and upgrade. Competition therefore provides perverse incentives for mercury reduction. This intuition is mapped in Table 5, but future research could formally model these negative impacts of competition in this interesting case of two civil society-backed programs (the typical dyad in the fragmentation literature is one civil society- versus one industry-backed program). Qualitative research on how miners choose between programs would deepen this contribution, as would linkage of program unification barriers to theories of conflict resolution.

Table 5. Program Competition Threatens Mercury Reduction

	Basic	Ecological
ARM	**Medium reward; Low effort**	Medium-High reward; High effort
FLO	Low reward; Low effort	High reward; High effort

Second, programs should make themselves more attractive to producers and consumers to encourage uptake and discourage decertification. Interviews with ARM (2018) revealed that some decertifications were driven by internal governance issues or natural disasters (e.g. mine collapse), but most were caused by insufficient earnings driven in part by low consumer demand. Increasing financial rewards to miners might therefore help. Currently 15% of the price of a certified gold wedding ring goes to the miner while 60% goes to the retailer (Author 2015b). While retailers have high expenses such as rent and advertising, changes such as moving stores online, becoming a benefit corporation to reduce taxes, and improving customer education to fetch higher prices could shift more money to miners with minimal changes to margins. Of course, findings from this case show that higher incomes may not lead to better behavior, and financial rewards can backfire by "crowding out" intrinsic motivations for change (Gneezy and Rustichini 2000). So non-monetary program enhancements should be explored, too.

Academics can support these efforts through research on miners' certification experiences and preferences. Such research should specifically focus on mercury-related issues (central to many SDGs) and related program changes might include: altering the reward differentials between basic and ecological certifications, requiring that some social premium be spent on mercury-free investments, requiring cost-sharing programs, forging partnerships with makers of mercury-free technology (e.g. businesses, universities), and offering miners education on alternative livelihoods.

Research should also focus on consumer demand. Certification programs prefer to write standards for products with *non*-luxury status (Author 2019) because a growing stream of research suggests an "attitude-behavior gap" exists for luxury consumption such that pro-social desires (e.g. ethical production) are over-ridden by self-serving desires (e.g. aesthetics) (Moraes et al. 2012). This is clearly a problem for certified gold. While consumers are willing to pay more for lower-priced items like certified coffee (Hainmueller et al. 2015), it is an open question how high up the price ladder pro-social behavior persists. And consumers complained to the joint program about gold's high price: for a typical wedding ring certification raised prices from roughly $558 to $754, nearly $200. Future research should rigorously test theories of luxury consumption to generate insights applicable to gold jewelry and other problematic luxury sector cases (e.g. furs, wine, cannabis, diamonds[2]).

Third, programs should consider which position within the ASGM regime to occupy. Programs could become dominant governors, weak but wide governors, strong but niche governors, or members of public-private partnerships (Auld et al. 2009). The presence of the Minamata Convention makes a dominant position unlikely. Success as a weak but wide governor is unlikely too, since programs could relax certain rules (e.g. paperwork requirements) to increase uptake, but weakening others (e.g. mercury requirements) would render programs illegitimate. More likely is success as a strong but niche governor: programs would *only* offer ecological certification, and certified miners would serve as living proof that best practices are possible. Success could also come via partnership with parties to the Minamata Convention or states working aggressively to achieve the SDGs, since public-private partnerships have a track record of success (Andonova 2017).

Alternatively, programs could decide that their best move is to disband. Programs likely struggle for goods reasons, and it is possible that certification and ASGM are not great institutional fits (Young 2011). Author (2019) identifies six sector characteristics that certification programs believe make products a good fit. Of these, marine aquarium fish had none and failed, while gold has only one and is struggling. Rather than artificially propping up struggling programs or working very hard to change them, the international community's time, energy and finances might be better invested in alternative governance approaches. Disbanded programs could join the Artisanal Gold Council's or Minamata Convention's existing capacity-building initiatives, focusing regimes' energies.

Whichever path programs choose, cautious optimism about certification's ability to contribute to the SDGs is warranted. Analysis of the gold case highlights both points of struggle and success, providing empirical ground for enhancing theories of certification's diffusion across product sectors and sustainability goals. As more research on sectors beyond agriculture, forestry and fisheries accrues, the conditions under which certification programs make the strongest contributions to regimes will be ascertained, to the benefit of the environment and international community. With wise evolution and discerning deployment in promising sectors, certification can be an important component of a suite of governance approaches that foster sustainable development.

References

Alliance for Responsible Mining (ARM), 2019. Community profile. http://www.fairmined.org/community-profiles-2/iquira/

Alliance for Responsible Mining (ARM). 2014. Fairmined Standard. Version 2.0.

Alliance for Responsible Mining (ARM). 2017. http://www.fairmined.org/our-impact/

Andonova, L. 2017. *Governance Entrepreneurs: International Organizations and the Rise of Global Public-Private.* Cambridge: Cambridge University Press.

[2] While consumers responded to 'blood diamonds,' problems in the sector remain (Smillie 2014).

Artisanal Gold Council (AGC), 2013. http://www.artisanalgold.org/publications/articles/mercury-recycling-in-artisanal-gold-mining-the-good-and-the-bad/

Auld, G., 2014. *Constructing Private Governance: The Rise and Evolution of Forest, Coffee, and Fisheries Certification*. Yale University Press.

Auld, G., Balboa C, Bernstein S, Cashore B. 2009. The emergence of non-state market-driven (NSMD) global environmental governance: a cross-sectoral assessment. In *Governance for the Environment: New Perspectives,* edited by Delmas, M and Young, O.

Auld G, Betsill M, VanDeveer S. 2018. Transnational Governance along the Mineral Lifecycle. *Annual Review of Environment and Resources*.

Auld G, Gulbrandsen LH, McDermott C. 2008. Certification Schemes and the Impact on Forests and Forestry. *Annual Review of Environment and Resources* 33:187-211

Bennett, E.A., 2018. Extending ethical consumerism theory to semi-legal sectors: insights from recreational cannabis. *Agriculture and human values*, pp.1-23.

Bennett, E.A., 2018(a). Voluntary Sustainability Standards: A Squandered Opportunity to Improve Workers' Wages. *Sustainable Development*, *26*(1), pp.65-82.

Bernstein, S., 2004. Legitimacy in global environmental governance. *J. Int'l L & Int'l Rel.*, *1*, p.139.

Biermann, Frank, Philipp Pattberg, Harro Van Asselt, and Fariborz Zelli. 2009. "The Fragmentation of Global Governance Architectures: A Framework for Analysis." *Global Environmental Politics* 9(4): 14–40.

Bloomfield M.J. 2015. *Dirty gold: How activism transformed the jewelry industry*. Cambridge, MA: MIT Press.

Bloomfield, M.J., & Schleifer, P. 2017. Tracing Failure of Coral Reef Protection in Nonstate Market-Driven Governance. *Global Environmental Politics*, 17(4), 127-146.

Cashore, B., Auld, G., Bernstein, S. and McDermott, C., 2007. Can non-state governance 'ratchet up'global environmental standards? lessons from the forest sector. *Review of European Community & International Environmental Law*, *16*(2), pp.158-172.

Cooley, A. and Ron, J., 2002. The NGO scramble: Organizational insecurity and the political economy of transnational action. *International security*, *27*(1), pp.5-39.

Childs, J., 2014. A new means of governing artisanal and small-scale mining? Fairtrade gold and development in Tanzania. *Resources Policy*, *40*, pp.128-136.

De Pena, 2015. "Taking Time to Reflect." https://www.huffingtonpost.co.uk/marike-de-peaa/fair-trade-sdgs_b_8291866.html?guccounter=1

Fisher, E., 2017. Solidarities at a distance: Extending Fairtrade gold to east Africa. *The Extractive Industries and Society*.

FLOCERT, 2017. Customer Database. https://www.flocert.net/about-flocert/customer-search/ Accessed December 2017.

Fairtrade International (FLO). 2013. Fairtrade Gold Standard.

Forest Stewardship Council (FSC), 2019. A Tool to Implement the SDGs. https://ic.fsc.org/file-download.fscr-a-tool-to-implement-the-sustainable-development-goals.a-1127.pdf

Gneezy, U. and Rustichini, A., 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), pp.791-810.

Green, J.F. and Auld, G., 2017. Unbundling the Regime Complex: The Effects of Private Authority. Transnational Environmental Law 6, 259–284.

Gulbrandsen, L. H. (2009). The emergence and effectiveness of the Marine Stewardship Council. *Marine Policy*, *33*(4), 654-660.

Hainmueller, J., Hiscox, M.J. and Sequeira, S., 2015. Consumer demand for fair trade: Evidence from a multistore field experiment. *Review of Economics and Statistics*, *97*(2), pp.242-256.

Hilson, G., 2008. 'Fair trade gold': Antecedents, prospects, challenges. *Geoforum*, *39*(1), pp.386-400.

Hilson, G., Hilson, A. and McQuilken, J., 2016. Ethical minerals: Fairer trade for whom?. *Resources Policy*, *49*, pp.232-247.

Hoffmann MJ. 2011. *Climate Governance at the Crossroads: Experimenting with a Global Response after Kyoto*. New York: Oxford University Press.

Ingle and Rhodes. 2016. https://www.ingleandrhode.co.uk/about-us/blog/why-choose-fairtrade-gold/. Accessed January 2018.

International Institutue for Sustainable Development (IISD), 2019. Standards and the Extractive Economy. IISD, Winnipeg, Canada.

Jaffee, D., 2014. *Brewing justice: Fair trade coffee, sustainability, and survival*. Univ of California Press.

Kalfagianni, A., & Pattberg, P. (2013). Fishing in muddy waters: Exploring the conditions for effective governance of fisheries and aquaculture. *Marine Policy*, *38*, 124-132.

Kanie, N. and Biermann, F., 2017. *Governing through goals: Sustainable development goals as governance innovation*. MIT Press.

Linton, A. 2012. *Fair Trade from the Ground Up: New Markets for Social Justice*. University of Washington Press.

Locke, R.M., 2013. *The promise and limits of private power: Promoting labor standards in a global economy*. Cambridge University Press.

Marine Stewardship Council (MSC), 2019. MSC and the UN SDGs. https://www.msc.org/about-the-msc/the-mscs-sustainability-goals

Moraes, Caroline, Marylyn Carrigan, and Isabelle Szmigin. "The coherence of inconsistencies: Attitude–behaviour gaps and new consumption communities." *Journal of Marketing Management* 28.1-2 (2012): 103-128.

Schleifer, P., Fiorini, M., Fransen, L. 2018. "Governing Tropical Agriculture through Voluntary Standards: A Mapping Analysis." Paper presented at ISA 59th Annual Convention, 4-7 April, 2018.

Selin, H. (2014). Global environmental law and treaty-making on hazardous substances: the Minamata Convention and mercury abatement. *Global Environmental Politics*, *14*(1), 1-19.

Selin, H., Keane, S.E., Wang, S., Selin, N.E., Davis, K. and Bally, D., 2018. Linking science and policy to support the implementation of the Minamata Convention on Mercury. *Ambio*, *47*(2), pp.198-215.

Short, J.L., Toffel, M.W. and Hugill, A.R., 2016. Monitoring global supply chains. *Strategic Management Journal*, *37*(9), pp.1878-1897.

Smillie, I., 2014. *Diamonds*, Cambridge: Polity.

Sun, Y. 2017. Transnational Public-Private Partnerships as Learning Facilitators: Global Governance of Mercury. *Global Environmental Politics*.

United Nations (UN), 2019. Sustainable Development Goals Knowledge Platform. Accessed 2019. https://sustainabledevelopment.un.org/sdgs

UN Environment Program (UNEP), 2019. Global Mercury Assessment 2018. UN Environment Programme, Chemicals and Health Branch. Geneva, Switzerland.

UN Forum on Sustainability Standards (UNFSS), 2018. Third Flagship Report: Voluntary Sustainability Standards, Trade, and Sustainable Development. Geneva, Switzerland. https://unfss.org/vss-and-the-sustainable-development-goals/

Vanderhaegen, K., Akoyi, K. T., Dekoninck, W., Jocqué, R., Muys, B., Verbist, B., & Maertens, M. (2018). Do private coffee standards 'walk the talk'in improving socio-economic and environmental sustainability?. *Global Environmental Change*, *51*, 1-9.

van der Ven, H., & Cashore, B. (2018). Forest certification: the challenge of measuring impacts. *Current Opinion in Environmental Sustainability*, *32*, 104-111.

van der Ven, H. C. Rothacker, and B. Cashore. 2018. Does Non-State Market-Driven Governance Create Unintended Land Use Impacts? Lessons from Fisheries, Palm Oil, Sugar, Cocoa and Forestry. Unpublished manuscript.

World Bank, 2018. Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle. Washington, DC: World Bank.

Young, O. R., 2011. Effectiveness of international environmental regimes: Existing knowledge, cutting-edge themes, and research strategies. *Proceedings of the National Academy of Sciences*, 108(50), 19853–19860

Zelli, F. and Van Asselt, H., 2013. Introduction: The institutional fragmentation of global environmental governance: Causes, consequences, and responses. *Global Environmental Politics*, *13*(3), pp.1-13.