RECONSTRUCTING ARRHENIUS'S IMPOSSIBILITY THEOREMS

TERUJI THOMAS

1. INTRODUCTION

Gustaf Arrhenius has developed a series of 'impossibility theorems' in population ethics, aiming to show that there can be no population axiology satisfying various combinations of plausible conditions.[1] These results are undoubtedly important, but the notation and the proofs themselves can be formidable. This note provides a much more concise (some might say, *too* concise) reconstruction of the proofs, using a notation which, I hope, will be more generally accessible. This also allows me to correct what appears to be a problem with his sixth and favoured theorem.

Here is one presentational choice I have made. I will present theorems in the form, 'Such-and-such conditions imply so-and-so variant of the Repugnant Conclusion'. Arrhenius, in contrast, states that such-and-such conditions are incompatible with so-and-so variant of what he calls 'The Quality Condition'. The Quality Condition entails the negation of the Repugnant Conclusion.

Now for the notation. We will restrict attention to a finite, linearly ordered set $\mathbb{W}$ of lifetime welfare levels, which I will represent by consecutive integers ranging from Z, which is far below zero, up to $A+2$, which is far above:

$$Z < Z+1 < \cdots < -1 < 0 < 1 < 2 < 3 < \cdots < A < A+1 < A+2.$$

Here 0 corresponds to a 'neutral' life, neither good nor bad for the person who lives it. So too -1 corresponds to a life with slightly negative welfare, and $1, 2, 3$ to lives with slightly positive welfare, or, in a common parlance, lives that are 'barely worth living'. Meanwhile, Z corresponds to a truly abominable life, and levels A and above correspond to truly excellent ones. The worse Z is, and the better A is, the more compelling will be the premisses of the theorems. But for them to be as compelling as possible, we must also suppose that this sequence of welfare levels is *fine grained*.[2] What this means is that the difference between consecutive levels is intuitively small, akin (for example) to the welfare gained by consuming a single gram of chocolate. Note the slight tension between the fact that there are finitely many welfare levels in the sequence between Z and A and the assumption that the sequence is fine-grained.

Arrhenius deals in *populations*, while I will deal in *welfare distributions*. A population is (Arrhenius says) a set of lives, whereas (I say) the corresponding welfare distribution is the number of lives at each welfare level. Formally, I understand a distribution to be a function from the set $\mathbb{W}$ of welfare levels to the set of non-negative integers. The value of the function at a given welfare level is the number of lives at that level. Arrhenius makes claims like this: a population with welfare distribution P is at least as good as a population with welfare distribution Q, all else

Date: April 2016.

[1]Arrhenius (2013). All of the theorems were published previously in some version. The first four theorems are essentially those developed in Arrhenius (2000a); the fifth is from Arrhenius (2003); the sixth is from Arrhenius (2009, 2011).

[2]This is Arrhenius's 'Discreteness' condition, or, more precisely, his claim that one can assume Discreteness without loss of generality.

equal. The 'all else equal' is necessary because populations determine non-welfare facts which may be axiologically relevant. But, abstracting away those non-welfare facts, it is harmless to simply say that P is at least as good as Q, and write $P \succsim Q$. I will assume that this betterness relation $\succsim$ is transitive and reflexive.

Given a welfare level w, I write $[\![w]\!]$ for the one-life distribution in which the one life has welfare w. I will write $\mathbb{N}$ for the set of natural numbers, starting from 1. Thus for $w \in \mathbb{W}$ and $n \in \mathbb{N}$, $n[\![w]\!]$ is the welfare distribution of a population of n lives, every one at level w. For example, the inequality

$$m[\![x]\!] + n[\![y]\!] \succsim N[\![v]\!]$$

states (in Arrhenius's terms) that a population of m people with welfare x and n people with welfare y is at least as good as a population of N people with welfare v , all else equal.

The table below may be useful to refer back to, later on. It lays out the conditions involved in each of Arrhenius's impossibility theorems. (I will use slightly weaker forms of some of his conditions.) The conditions are groups thematically. Annotations in the left-most column indicate some of the logical implications between the conditions. Some hypotheses are invariably used to derive intermediary conditions, which I include below them in italics.

			Theorems					
Abbreviations		**Hypotheses**	**1**	**2**	**3**	**4**	**5**	**6**
	ED	Egalitarian Dominance	X	X	X	X	X	X
	Q	Quantity	X					
	DA	Dominance Addition		X			X	
	NS	Non-Sadism			X			
NS $\Longrightarrow$	WNS	Weak Non-Sadism				X		X
	GNE	General Non-Elitism					X	
GNE $\Longrightarrow$	*GIAA*	*Gen. Inequality-Averse Addition* (β')						
GNE $\Longrightarrow$	NE	Non-Elitism (almost the same!)				X		X
NE $\Longrightarrow$	*IAA*	*Inequality-Averse Addition* (β)						
IAA $\Longrightarrow$	IA	Inequality Aversion		X	X			
	GNEP	General Non-Extreme Priority				X	X	X
GNEP $\Longrightarrow$	*ST*	*Sufficient Tradeoffs* (δ)						
GNEP $\Longrightarrow$	NEP	Non-Extreme Priority			X			
		Conclusions						
	RC	Repugnant Conclusion[3]	X	X				
RC $\Longrightarrow$	RA	Repugnant Addition[4]			X	X		
RC $\Longleftarrow$	VRC	Very Repugnant Conclusion[5]					X	
VRC $\Longrightarrow$	VRA*	Very Repugnant Addition*[6]						X

[3]Contradicts the 'Quality Condition'.

[4]Contradicts the 'Quality Addition Condition'.

[5]Contradicts the 'Weak Quality Condition'.

[6]VRA contradicts the 'Weak Quality Addition Condition'; VRA* is a variant – see comments below.

2. The First Impossibility Theorem

The first adequacy condition we will consider is a very weak version of the principle that increasing welfare levels increases a population's value. It is weak because it only considers populations in which everyone is equally well off.

(ED) Egalitarian Dominance. For any welfare levels $x < y$ and any population size $n \in \mathbb{N}$,

$$n[\![x]\!] \prec n[\![y]\!].$$

This condition is illustrated in Figure 1, with the usual convention that the width of a box is indicative of the size of a group of people, and the height is indicative of their welfare level. In all these figures, the distribution on the right is claimed to be at least as good as the one on the left.[7]

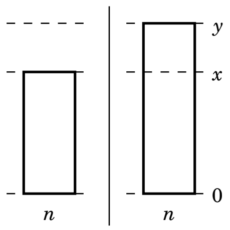

FIGURE 1. Egalitarian Dominance (ED)

The second adequacy condition is one of several to the effect that adding positive-welfare lives to a population does not make it worse. This one says, more specifically, that a small decrease in welfare levels can be compensated by increasing the size of the population.

(Q) Quantity. For any welfare level $w > 1$ and any $m \in \mathbb{N}$ there exists $Q(w,m) \in \mathbb{N}$ such that

$$m[\![w]\!] \precsim Q(w,m)[\![w-1]\!].$$

See Figure 2. Note that Quantity gains plausibility from the assumption that the sequence of welfare levels is fine-grained, so that the decrease in welfare levels is small. It is less intuitive that a *large* decrease in welfare levels can be made up by an increase in population size. Indeed, that would immediately entail the Repugnant Conclusion, which in the current framework looks like this:

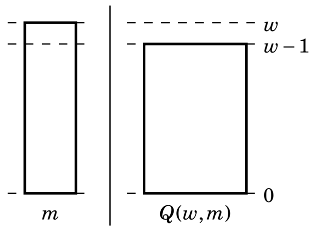

FIGURE 2. Quantity (Q)

(RC) Repugnant Conclusion. For any $m \in \mathbb{N}$, there exists $M(m) \in \mathbb{N}$ such that

$$m[\![A]\!] \prec M[\![3]\!].$$

[7] I thank Daniel Ramöller for his online box-diagram creator boxethics.org.

See Figure 3.

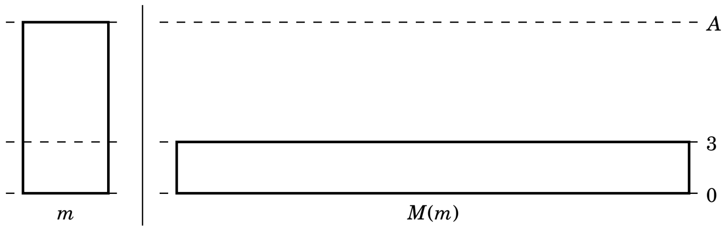

FIGURE 3. Repugnant Conclusion (RC)

Theorem 1. *ED and Q jointly entail RC.*

Proof. Fix any m as in RC. Use Quantity to define $Q_0 = Q(A,m)$, and then, for i from 1 up to $A-1$, recursively define $Q_i = Q(A-i, Q_{i-1})$. Then, applying Quantity $A-3$ times, we have

$$m[\![A]\!] \precsim Q_0[\![A-1]\!] \precsim Q_1[\![A-2]\!] \precsim \cdots \precsim Q_{A-3}[\![2]\!].$$

Since, by ED, $Q_{A-3}[\![2]\!] \prec Q_{A-3}[\![3]\!]$, we have RC with $M(m) := Q_{A-3}$. □

3. THE SECOND IMPOSSIBILITY THEOREM

The next adequacy condition is another version of the idea that adding positive-welfare lives does not make a population worse. Roughly speaking, it says that if we add positive-welfare lives and improve the pre-existing lives, the resulting population is at least as good as the original.

(DA) Dominance Addition. For any welfare levels $x > y$ and $z > 0$, and any $m, n \in \mathbb{N}$,

$$m[\![y]\!] \precsim m[\![x]\!] + n[\![z]\!].$$

See Figure 4. Now, there is a technicality here. Arrhenius usually states DA as claiming only that the left-hand side is *not better* than the right-hand side. On the other hand, he (superficially) uses my stronger version in proving his fifth theorem. I will use the stronger version, but will remark after Theorems 2 and 5 on the very minor changes needed to accommodate the weaker version.

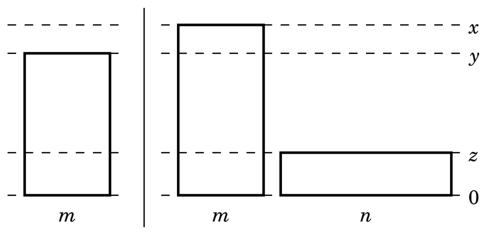

FIGURE 4. Dominance Addition (DA)

The next adequacy condition is meant to be a necessary condition for an axiology to be egalitarian, or at least not strongly anti-egalitarian.

(IA) Inequality Aversion. For any $x > y > z \in \mathbb{W}$ and any $m \in \mathbb{N}$, there exists $C(x,y,z,m) \in \mathbb{N}$ such that

$$m[\![x]\!] + C[\![z]\!] \precsim m[\![y]\!] + C[\![y]\!].$$

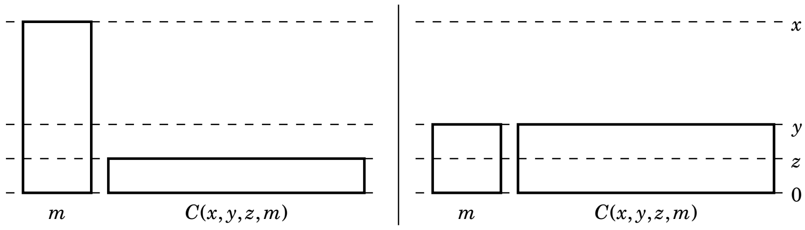

FIGURE 5. Inequality Aversion (IA)

See Figure 5.

I claim that IA, like Quantity, gains some of its plausibility from the assumption that the sequence of welfare levels is fine-grained. If the value difference between x and y could be infinitely or categorically more important than the value difference between y and z, then violations of IA would be unsurprising. A few losses of welfare from x to y might outweigh any number of gains from z to y. But fine-grainedness says that we can get between x and y, and between y and z, in a finite number of 'small' steps. This appears to rule out such categorical differences. In fact, the fourth, fifth, and sixth theorems use this line of thought to argue for variations of IA from 'Non-Elitism'.

Theorem 2. *ED, DA, and IA jointly entail RC.*

Proof. Fix $m \in \mathbb{N}$. For $C \in \mathbb{N}$ to be determined as below, we have:

$$\begin{aligned} m[\![A]\!] &\precsim m[\![A+1]\!] + C[\![1]\!] && \text{(Dominance Addition)} \\ &\precsim m[\![2]\!] \quad + C[\![2]\!] && \text{(Inequality Aversion gives } C(A+1,2,1,m)) \\ &\prec m[\![3]\!] \quad + C[\![3]\!] && \text{(Egalitarian Dominance)} \end{aligned}$$

In summary, we have found $M := m + C$ with $m[\![A]\!] \prec M[\![3]\!]$. □

Now, if we adopt the weaker version of DA, then the valid conclusion of the argument is that $m[\![A]\!]$ is *not at least as good as* $M[\![3]\!]$. This is still a violation of Arrhenius's 'Quality Condition', but it is a little weaker than what is required for RC (namely, that $m[\![A]\!]$ be *better than* $M[\![3]\!]$).

4. THE THIRD IMPOSSIBILITY THEOREM

The next condition expresses another variation on the idea that adding positive-welfare lives cannot make things worse. More precisely, it says that adding any number of positive-welfare lives is at least as good as adding any number of negative-welfare lives.

(NS) Non-Sadism. For any welfare levels $x > 0 > z$, any $m, n \in \mathbb{N}$, and any distribution P,

$$P + n[\![z]\!] \precsim P + m[\![x]\!].$$

See Figure 6.

The next condition expresses the idea that raising welfare just across the netural level is not categorically more important than raising welfare from very low positive to very high positive levels.

(NEP) Non-Extreme Priority. There exists $B \in \mathbb{N}$ such that for any distribution P,

$$\begin{aligned} &P + [\![3]\!] \quad + B[\![3]\!] \\ \precsim\ &P + [\![-1]\!] + B[\![A]\!]. \end{aligned}$$

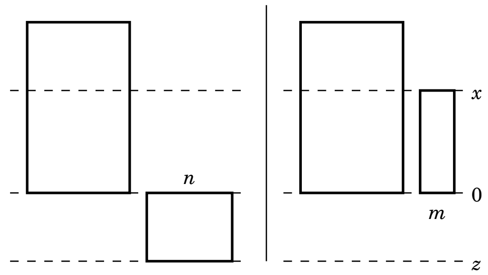

FIGURE 6. Non-Sadism (NS)

See Figure 7.

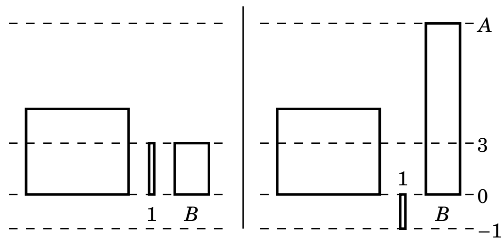

FIGURE 7. Non-Extreme Priority (NEP)

The third theorem will derive a version of RC which, like NS and NEP, includes an unaffected background distribution P.

(RA) Repugnant Addition. For some distribution P, and any $n \in \mathbb{N}$, there exists $N \in \mathbb{N}$ such that

$$P + n[\![A]\!] \prec P + N[\![3]\!].$$

Theorem 3. *ED, IA, NS, and NEP jointly entail RA.*

Proof. Fix any $n \in \mathbb{N}$. For $B, C \in \mathbb{N}$ to be determined as below, we have

$$\begin{aligned}
& [\![3]\!] + B[\![3]\!] + n[\![A]\!] \\
\precsim\ & [\![-1]\!] + B[\![A]\!] + n[\![A]\!] && \text{(NEP gives } B\text{).} \\
\precsim\ & C[\![1]\!] + B[\![A]\!] + n[\![A]\!] && \text{(Non-Sadism)} \\
\prec\ & C[\![2]\!] + B[\![2]\!] + n[\![2]\!] && \text{(Inequality Aversion gives } C(A,2,1,B+n)\text{)} \\
\precsim\ & C[\![3]\!] + B[\![3]\!] + n[\![3]\!] && \text{(Egalitarian Dominance).}
\end{aligned}$$

We can rewrite the overall inequality as

$$(1+B)[\![3]\!] + n[\![A]\!] \prec (1+B)[\![3]\!] + (C+n-1)[\![3]\!].$$

This implies Repugnant Addition for $P := (1+B)[\![3]\!]$ and $N := C+n-1$. Note that P is independent of n, as required by RA. □

5. LEMMA: NON-ELITISM AND INEQUALITY-AVERSE ADDITION

The fourth, fifth, and sixth theorems derive a generalisations of Inequality Aversion from 'Non-Elitism' (Arrhenius's Lemmas 4.1 and 5.1). Arrhenius presents two versions of Non-Elitism that differ only slightly, and the argument is really the same in each case.

(GNE) General Non-Elitism. For all $x, z \in \mathbb{W}$, with $x > z+1$, there exists $G(x,z) \in \mathbb{N}$ such that, for any population R,

$$\begin{aligned} & R + [\![x]\!] \quad\;\; + G(x,z)[\![z]\!] \\ \precsim\; & R + [\![x-1]\!] + G(x,z)[\![x-1]\!]. \end{aligned}$$

See Figure 8. The rough idea of General Non-Elitism is that increasing the welfare

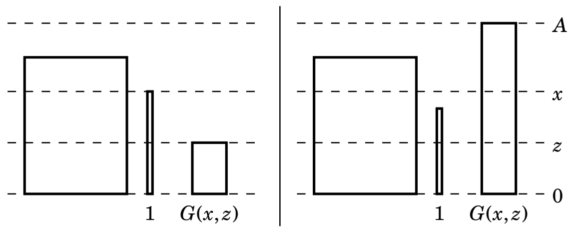

FIGURE 8. General Non-Elitism (GNE)

of one person from $x-1$ to x cannot be categorically more important than increasing the welfare of worse-off people from z to $x-1$. The plausibility of General Non-Elitism depends on the idea that the value difference between x and $x-1$ is not categorically greater than the value difference between $x-1$ and z. Thus, like IA and Q, its plausibility depends to some extent on the assumption that the sequence of welfare levels is fine-grained.

General Non-Elitism is used to derive the following variant on IA, which include an unaffected background distribution P.

(GIAA) General Inequality-Averse Addition.[8] For any $x > y > z \in \mathbb{W}$ and any $m \in \mathbb{N}$, there exists $C(x,y,z,m) \in \mathbb{N}$ such that, for any population P,

$$\begin{aligned} & P + m[\![x]\!] + C[\![z]\!] \\ \precsim\; & P + m[\![y]\!] + C[\![y]\!]. \end{aligned}$$

Arrhenius's Lemma 4.1 uses a restricted (not 'General') version of **Non-Elitism (NE)**. This differs from the 'General' one above only in that it restricts the welfare levels occurring in R to lie between z and x. The corresponding variant on IA, which I will call simply **Inequality-Averse Addition (IAA)**,[9] restricts the welfare levels occurring in P to lie between z and $y+1$.

Lemma 1. *GNE entails GIAA. (Similarly NE entails IAA.)*

Proof. Fix any x, y, z, m as in GIAA. Set $m_1 = m$ and $C_1 = mG(x,z)$. Then, for $i = 2$ up to $i = x - y$, define recursively

$$m_i = m_{i-1} + C_{i-1} \qquad\qquad C_i = m_i G(x-i+1, z).$$

We will derive GIAA with $C(x,y,z,m) = C_1 + C_2 + \cdots + C_{x-y}$. Indeed, we obtain the following inequalities by applying GNE first m_1 times, then m_2 times, and so on,

[8] Arrhenius's Condition β'.

[9] Arrhenius's Condition β.

until finally we apply it m_{x-y} times:

$$\begin{aligned}
&P &&+m[\![x]\!] &&+C_1[\![z]\!] &&+C_2[\![z]\!] &&+\cdots+C_{x-y}[\![z]\!]\\
\precsim\, &P &&+m[\![x-1]\!] &&+C_1[\![x-1]\!] &&+C_2[\![z]\!] &&+\cdots+C_{x-y}[\![z]\!]\\
\precsim\, &P &&+m[\![x-2]\!] &&+C_1[\![x-2]\!] &&+C_2[\![x-2]\!] &&+\cdots+C_{x-y}[\![z]\!]\\
\precsim\, &\cdots\cdots\cdots\\
\precsim\, &P &&+m[\![y]\!] &&+C_1[\![y]\!] &&+C_2[\![y]\!] &&+\cdots+C_{x-y}[\![y]\!]
\end{aligned}$$

Overall we have $P+m[\![x]\!]+C[\![z]\!]\precsim P+m[\![y]\!]+C[\![y]\!]$, as desired. □

6. Lemma: General Non-Extreme Priority and Sufficient Tradeoffs

The last three theorems require a further lemma, based on the following adequacy condition. (This corresponds to Lemma 4.2 in Arrhenius.) It expresses the idea that increasing one person's welfare by a small amount is never categorically more important than raising the welfare of others from very low positive to very high positive levels.

(GNEP) General Non-Extreme Priority. For any $z \in \mathbb{W}$, there exists $G(z) \in \mathbb{N}$ such that, for any welfare level $x \geq A$, any welfare level $0 < y \leq 3$, and any population P,

$$\begin{aligned}
&P+[\![z]\!] &&+G(z)[\![y]\!]\\
\precsim\, &P+[\![z-1]\!] &&+G(z)[\![x]\!].
\end{aligned}$$

See Figure 9. Once more, GNEP gains plausibility from the assumption that the

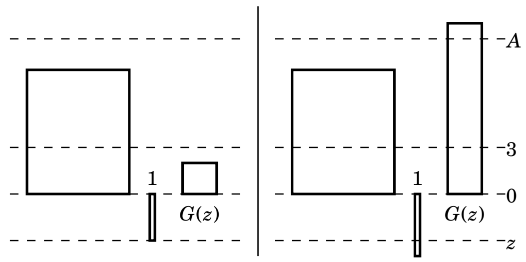

Figure 9. General Non-Extreme Priority (GNEP)

sequence of welfare levels is fine-grained, so that the difference between z and $z-1$ is always small.[10]

GNEP is used to derive the following more general condition on interpersonal tradeoffs.

(ST) Sufficient Tradeoffs.[11] For any welfare levels $z < 0 < y \leq 3$ and any $m \in \mathbb{N}$, there is some $B(z,m) \in \mathbb{N}$ such that, for any welfare level $x \geq A$ and any population P,

$$\begin{aligned}
&P+m[\![y]\!]+B(z,m)[\![y]\!]\\
\precsim\, &P+m[\![z]\!]+B(z,m)[\![x]\!].
\end{aligned}$$

See Figure 10.

Roughly speaking, raising some people's welfare from negative levels to very low positive levels is never categorically more important than raising other people's welfare from very low positive to very high positive levels.

[10]In some but not all of Arrhenius's 'informal' versions of GNEP, G depends on P. But then the proofs won't work.

[11]Arrhenius's Condition δ; also called 'bad lives for very good lives' in his §6.8.

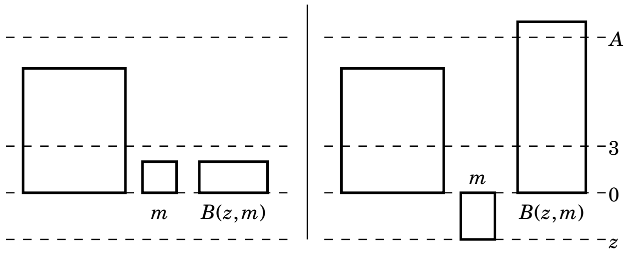

FIGURE 10. Sufficient Tradeoffs (ST)

Lemma 2. *GNEP entails ST.*

Proof. Fix any z, y, m as in ST. For $i = 1$ up to $i = y - z$, use GNEP to define $B_i = mG(z+i)$. We will derive ST with $B(z,m) := B_1 + B_2 + \cdots + B_{3-z}$. Indeed, for any x, P as in ST, we obtain the following inequalities, applying GNEP m times at each stage:

$$\begin{aligned}
& P && + m[\![y]\!] + B_1[\![y]\!] + B_2[\![y]\!] + \cdots + B_{y-z}[\![y]\!] \\
\precsim\ & P && + m[\![2]\!] + B_1[\![x]\!] + B_2[\![y]\!] + \cdots + B_{y-z}[\![y]\!] \\
\precsim\ & P && + m[\![1]\!] + B_1[\![x]\!] + B_2[\![x]\!] + \cdots + B_{y-z}[\![y]\!] \\
\precsim\ & \cdots\cdots\cdots \\
\precsim\ & P && + m[\![z]\!] + B_1[\![x]\!] + B_2[\![x]\!] + \cdots + B_{y-z}[\![x]\!]
\end{aligned}$$

Overall we have $P + m[\![y]\!] + B[\![y]\!] \precsim P + m[\![z]\!] + B[\![x]\!]$, as desired. □

7. FOURTH IMPOSSIBILITY THEOREM

The fourth impossibility theorem involves a notional weakening of Non-Sadism. It says, roughly, that adding any number of positive-welfare lives is at least as good as adding a very large number of abominable ones.

(WNS) Weak Non-Sadism. There is some $D \in \mathbb{N}$, such that, for any welfare level $x > 0$, any $m \in \mathbb{N}$, and any population P,

$$P + D[\![Z]\!] \precsim P + m[\![x]\!].$$

See Figure 11.

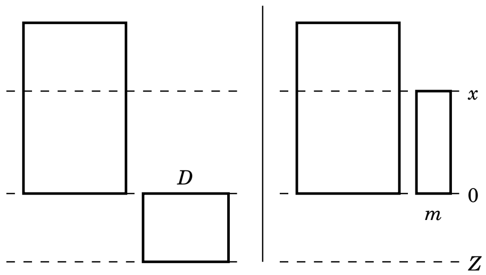

FIGURE 11. Weak Non-Sadism (WNS)

Theorem 4. *ED, GNEP, NE, and WNS jointly entail Repugnant Addition.*

Proof. For given $n \in \mathbb{N}$, and $B, C, D, F \in \mathbb{N}$ to be determined:

$$\begin{array}{lll} & n[\![A]\!] + B[\![3]\!] + 2D[\![3]\!] & \\ \precsim & n[\![A]\!] + B[\![A]\!] + 2D[\![Z]\!] & \text{ST/GNEP gives } B(Z, 2D) \\ \precsim & n[\![A]\!] + B[\![A]\!] + C[\![1]\!] + F[\![2]\!] & \text{WNS (twice) gives } D \\ \precsim & n[\![2]\!] + B[\![2]\!] + C[\![2]\!] + F[\![2]\!] & \text{IAA/NE gives } C(A, 2, 1, n+B) \\ \prec & n[\![3]\!] + B[\![3]\!] + C[\![3]\!] + F[\![3]\!] & \text{Egalitarian Dominance.} \end{array}$$

For any $F > 2D$ we can rewrite the overall comparison as

$$(B+2D)[\![3]\!] + n[\![A]\!] \prec (B+2D)[\![3]\!] + (n+C+F-2D)[\![3]\!].$$

This entails Repugnant Addition, with $P := (B+2D)[\![3]\!]$ and $N := n+C+F-2D$. Note that, as required, P does not implicitly depend on n. □

8. Fifth Impossibility Theorem

The fifth theorem derives the following strong version of RC.

(VRC) Very Repugnant Conclusion. For any $m, n \in \mathbb{N}$, and any welfare level $z < 0$, there exists $N(m,n,z) \in \mathbb{N}$ such that

$$m[\![z]\!] + N[3] \succ n[\![A]\!].$$

See Figure 12.

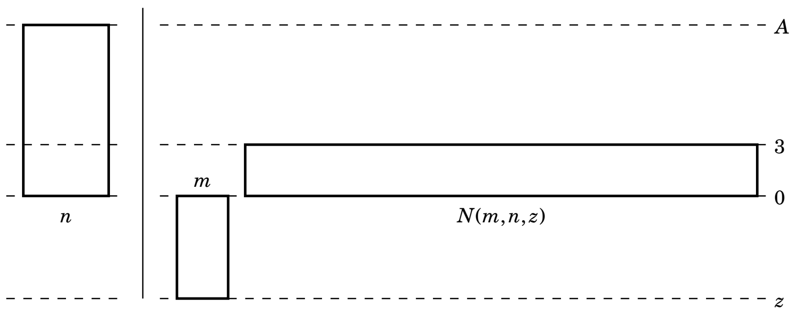

Figure 12. The Very Repugnant Conclusion (VRC)

Theorem 5. *DA, ED, GNE, and GNEP jointly entail VRC.*

Proof. For given $m, n \in \mathbb{N}$, given $z < 0$, and B, C to be determined:

$$\begin{array}{lll} & n[\![A]\!] & \\ \prec & n[\![A+1]\!] & \text{Egalitarian Dominance} \\ \precsim & n[\![A+2]\!] + m[\![1]\!] + B[\![1]\!] + C[\![1]\!] & \text{Dominance Addition} \\ \precsim & n[\![A+2]\!] + m[\![z]\!] + B[\![A+2]\!] + C[\![1]\!] & \text{ST/GNEP gives } B(z, m) \\ \precsim & n[\![3]\!] + m[\![z]\!] + B[\![3]\!] + C[\![3]\!] & \text{GIAA/GNE gives } C(A+2, 3, 1, n+B) \\ = & m[\![z]\!] + n[\![3]\!] + B[\![3]\!] + C[\![3]\!]. & \end{array}$$

Thus we obtain VRC with $N := n + B + C$. □

Now, if we adopt the weaker version of DA, then the valid conclusion of the argument is that $n[\![A]\!]$ is *not at least as good as* $m[\![z]\!] + N[\![3]\!]$. This is still a violation of Arrhenius's 'Weak Quality Condition', although slightly weaker than what is required for VRC (namely, that $n[\![A]\!]$ be *better than* $m[\![z]\!] + N[\![3]\!]$).

9. Sixth Impossibility Theorem

Arrhenius's sixth theorem has the same premisses as the fourth, but claims to derive the following version of VRC which, like RA, includes an unaffected background population.

(VRA) Very Repugnant Addition. For some population P, any welfare level $z < 0$, and any $m, n \in \mathbb{N}$, there exists $N(m,n,z) \in \mathbb{N}$ such that

$$P + m[\![z]\!] + N[\![3]\!] \succ P + n[\![A]\!].$$

However, there is a mistake in Arrhenius's proof. What can be proved, instead of VRA, is:

(VRA*) Very Repugnant Addition*. For any welfare level $z < 0$ and any $m \in \mathbb{N}$, there is some population P such that, for any $n \in \mathbb{N}$, there exists $N \in \mathbb{N}$ with

$$P + m[\![z]\!] + N[\![3]\!] \succ P + n[\![A]\!].$$

This is slightly weaker than VRA, because P is allowed to depend z and m. I think that VRA* is not much less 'repugnant' than VRA, but then again it is not easy to understand the difference between them.

Theorem 6. *ED, GNEP, NE, and WNS jointly entail VRA*.*

Proof. Recall that in proving the fourth theorem we obtained RA with N of the form $n+C+F-2D$, and that the only constraint on F was $F > 2D$. We can therefore choose that N to be larger than any given N_0. Renaming the variables for convenience, we have:

Restricted Repugnant Addition. For some population P', any $n' \in \mathbb{N}$, and any $N_0 \in \mathbb{N}$, there exists $N' > N_0 \in \mathbb{N}$ such that

$$P' + N'[\![3]\!] \succ P' + n'[\![A]\!].$$

I then claim that RRA and GNEP jointly entail VRA* (this is a corrected version of Arrhenius's Lemma 6.1). For z, m, n as in VRA*, P' as in RRA, and B, N to be determined:

$$\begin{aligned}
& P' + B[\![A]\!] + n[\![A]\!] && \\
\prec\ & P' + N'[\![3]\!] && \text{RRA for } n' = B + n \text{ determines } N' > B + m \\
=\ & P' + B[\![3]\!] + m[\![3]\!] + N[\![3]\!] && \text{Define } N = N' - B - m \\
\precsim\ & P' + B[\![A]\!] + m[\![z]\!] + N[\![3]\!] && \text{ST/GNEP determines } B(z,m).
\end{aligned}$$

We therefore get VRA* with background population $P := P' + B[\![A]\!]$. Note that the conclusion is VRA* rather than VRA because P depends on m and z through B. □

References

Arrhenius, Gustaf 2000a, *Future Generations: A Challenge for Moral Theory* (Uppsala: University Printers)

—— 2000b, 'An impossibility theorem for welfarist axiologies', in *Economics and Philosophy* 16

—— 2003, 'The very repugnant conclusion', in K. Segerberg and R. Sliwinski (eds.), *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Aqvist* (Uppsala: Uppsala University)

—— 2005, 'The paradoxes of future generations and normative theory', in J. Ryberg and T. Tännsjö (eds.), *The Repugnant Conclusion: Essays on Population Ethics* (Dordrecht: Kluwer Academic Publishers)

—— 2009, 'One more axiological impossibility theorem', in L. Johansson, J. Österberg, and R. Sliwinski (eds.), *Logic, Ethics, and All That Jazz: Essays in Honour of Jordan Howard Sobel* (Uppsala: Uppsala University)
—— 2011, 'The impossibility of a satisfactory population ethics', in E.N. Dzhafarov and L. Perry (eds.), *Descriptive and Normative Approaches to Human Behavior* (Singapore: World Scientific Publishing Co.)
—— 2013, *Population Ethics* (manuscript)