Extensionale Interpretation des Kategorischen Imperativs

Cornelius Diekmann

November 12, 2022

Contents

1 Schnelleinstieg Isabelle/HOL

1.1 Typen

Typen werden per :: annotiert. Beispielsweise sagt $3::nat$, dass 3 eine natürliche Zahl (nat) ist.

1.2 Beweise

Die besondere Fähigkeit im Beweisassistent Isabelle/HOL liegt darin, maschinengeprüfte Beweise zu machen.
Beispiel:

lemma‹$3 = 2+1$›

In der PDFversion wird der eigentliche Beweis ausgelassen. Aber keine Sorge, der Computer hat den Beweis überprüft. Würde der Beweis nicht gelten, würde das PDF garnicht compilieren.

Ich wurde schon für meine furchtbaren Beweise zitiert. Ist also ganz gut, dass wir nur Ergebnisse im PDF sehen und der eigentliche Beweis ausgelassen ist. Am besten kann man Beweise sowieso im Isabelle Editor anschauen und nicht im PDF.

1.3 Mehr Typen

Jeder Typ der mit einem einfachen Anführungszeichen anfängt ist ein polymorpher Typ. Beispiel: $'a$ oder $'\alpha$. So ein Typ ist praktisch ein generischer Typ, welcher durch jeden anderen Typen instanziiert werden kann.

Beispielsweise steht $'nat$ für einen beliebigen Typen, während nat der konkrete Typ der natürlichen Zahlen ist.

Wenn wir nun $3::'a$ schreiben handelt es sich nur um das generische Numeral 3. Das ist so generisch, dass z.B. noch nicht einmal die Plusoperation darauf definiert ist. Im Gegensatz dazu ist $3::nat$ die natürliche Zahl 3, mit allen wohlbekannten Rechenoperationen. Im Beweis obigen **lemma**‹$3 = 2+1$› hat Isabelle die Typen automatisch inferiert.

1.4 Funktionen

Beispiel: Eine Funktionen welche eine natürliche Zahl nimmt und eine natürliche Zahl zurück gibt ($nat \Rightarrow nat$):

fun *beispielfunktion* :: ‹$nat \Rightarrow nat$› **where**
‹*beispielfunktion* $n = n + 10$›

Funktionsaufrufe funktionieren ohne Klammern.

lemma ‹*beispielfunktion* $32 = 42$›

Funktionen sind gecurried. Hier ist eine Funktion welche 2 natürliche Zahlen nimmt und eine natürliche Zahl zurück gibt ($nat \Rightarrow nat \Rightarrow nat$):

fun *addieren* :: ‹*nat* ⇒ *nat* ⇒ *nat*› **where**
‹*addieren a b* = *a* + *b*›

lemma ‹*addieren 32 10* = *42*›

Currying bedeutet auch, wenn wir *addieren* nur mit einem Argument aufrufen (welches eine natürliche Zahl *nat* sein muss), dass wir eine Funktion zurückbekommen, die noch das zweite Argument erwartet, bevor sie das Ergebnis zurückgeben kann.
Beispiel: *addieren 10*::*nat* ⇒ *nat*
Zufälligerweise ist *addieren 10* equivalent zu *beispielfunktion*:

lemma ‹*addieren 10* = *beispielfunktion*›

Zusätzlich lassen sich Funktionen im Lambda Calculus darstellen. Beispiel:

lemma ‹(λ*n*::*nat*. *n*+*10*) *3* = *13*›

lemma ‹*beispielfunktion* = (λ*n*. *n*+*10*)›

1.5 Mengen

Mengen funktionieren wie normale mathematische Mengen.
Beispiel. Die Menge der geraden Zahlen:

lemma ‹{*0*,*2*,*4*,*6*,*8*,*10*,*12*} ⊆ {*n*::*int*. *n mod 2* = *0*}›

2 Disclaimer

Ich habe

- kein Ahnung von Philosophie.
- keine Ahnung von Recht und Jura.
- und schon gar keine Ahnung von Strafrecht oder Steuerrecht.

Und in dieser Session werden ich all das zusammenwerfen.

Cheers!

2.1 Über den Titel

Der Titel lautet *Extensionale Interpretation des Kategorischen Imperativs.* Dabei sind die Wörter wie folgt zu verstehen

- *Extensional* bezieht sich hier auf den Fachbegriff der Logik https://en.wikipedia.org/wiki/Extensionality, welcher besagt, dass Objekte gleich sind, wenn sie die gleichen externen Eigenschaften aufweisen. Beispielsweise sind zwei Funktionen gleich, wenn sie für alle Eingaben die gleiche Ausgabe liefern: $(f = g) = (\forall x.\ f\ x = g\ x)$. Die interne (intensionale) Implementierung der Funktionen mag unterschiedlich sein, dennoch sind sie gleich. Dies ist die natürliche Gleichheit in HOL, welche uns erlaubt unser Modell bequem zu shallow-embedden. Meine extensionale Modellierung prägt diese Theorie stark. Beispielsweise sind Handlungen extensional modelliert, d.h nur die äußerlich messbaren Ergebnisse werden betrachtet. Dies widerspricht vermutlich stark Kants Vorstellung.
- *Interpretation* besagt, dass es sic hier um meine persönliche Interpretation handelt. Diese Theorie ist keine strenge Formalisierung der Literatur, sondern enthält sehr viele persönliche Meinungen.
- *Kategorischer Imperativ* bezieht sich auf Kants Kategorischer Imperativ. Ziel dieser Theorie ist es, moralische Entscheidungen basierend auf Kants Idee zu machen.

3 Handlung

Beschreibt Handlungen als Änderung der Welt. Unabhängig von der handelnden Person. Wir beschreiben nur vergangene bzw. mögliche Handlungen und deren Auswirkung.

Eine Handlung ist reduziert auf deren Auswirkung. Intention oder Wollen ist nicht modelliert, da wir irgendwie die geistige Welt mit der physischen Welt verbinden müssen und wir daher nur messbare Tatsachen betrachten können.

Handlungen können Leute betreffen. Handlungen können aus Sicht Anderer wahrgenommen werden. Ich brauche nur Welt vorher und Welt nachher. So kann ich handelnde Person und beobachtende Person trennen.

datatype *'world handlung = Handlung (vorher: ‹'world›) (nachher: ‹'world›)*

definition *ist-noop* :: ‹*'world handlung ⇒ bool*› **where**
‹*ist-noop h ≡ vorher h = nachher h*›

Handlung als Funktion gewrapped. Diese abstrakte Art eine Handlung zu modelliert so ein bisschen die Absicht oder Intention.

datatype (*'person, 'world*) *handlungsabsicht = Handlungsabsicht ‹'person ⇒ 'world ⇒ 'world›*

Von Außen können wir Funktionen nur extensional betrachten, d.h. Eingabe und Ausgabe anschauen. Die Absicht die sich in einer Funktion verstecken kann ist schwer zu erkennen. Dies deckt sich ganz gut damit, dass Isabelle standardmäßig Funktionen nicht printet. Eine $('person, 'world)$ *handlungsabsicht* kann nicht geprinted werden!

fun *handeln* :: ‹*'person* ⇒ *'world* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ *'world handlung*› **where**
‹*handeln handelnde-person welt* (*Handlungsabsicht h*) = *Handlung welt* (*h handelnde-person welt*)›

Beispiel, für eine Welt die nur aus einer Zahl besteht: Wenn die Zahl kleiner als 9000 ist erhöhe ich sie, ansonsten bleibt sie unverändert.

definition ‹*beispiel-handlungsabsicht* ≡ *Handlungsabsicht* (λ- *n*. *if n* < *9000 then n*+*1 else n*)›

Da Funktionen nicht geprintet werden können, sieht *beispiel-handlungsabsicht* so aus: *Handlungsabsicht -*

3.1 Interpretation: Gesinnungsethik vs. Verantwortungethik

Sei eine Ethik eine Funktion, welche einem beliebigen $'\alpha$ eine Bewertung Gut = *True*, Schlecht = *False* zuordnet.

- Eine Ethik hat demnach den Typ: $'\alpha \Rightarrow$ *bool*.

Laut https://de.wikipedia.org/wiki/Gesinnungsethik ist eine Gesinnungsethik "[..] eine der moralischen Theorien, die Handlungen nach der Handlungsabsicht [...] bewertet, und zwar ungeachtet der nach erfolgter Handlung eingetretenen Handlungsfolgen."

- Demnach ist eine Gesinnungsethik: (*'person*, *'world*) *handlungsabsicht* ⇒ *bool*.

Nach https://de.wikipedia.org/wiki/Verantwortungsethik steht die Verantwortungsethik dazu im strikten Gegensatz, da die Verantwortungsethik "in der Bewertung des Handelns die Verantwortbarkeit der *tatsächlichen Ergebnisse* betont."

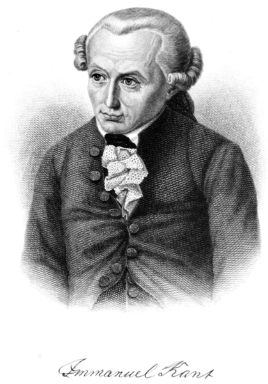
- Demnach ist eine Verantwortungsethik: *'world handlung* ⇒ *bool*.

Da *handeln* eine Handlungsabsicht (*'person*, *'world*) *handlungsabsicht* in eine konkrete Änderung der Welt *'world handlung* überführt, können wie die beiden Ethiktypen miteinander in Verbindung setzen. Wir sagen, eine Gesinnungsethik und eine Verantwortungsethik sind konsistent, genau dann wenn für jede Handlungsabsicht, die Gesinnungsethik die Handlungsabsicht genau so bewertet, wie die Verantwortungsethik die Handlungsabsicht bewerten würde, wenn die die Handlungsabsicht in jeder möglichen Welt und als jede mögliche handelnde Person tatsächlich ausführt wird und die Folgen betrachtet werden:

definition *gesinnungsethik-verantwortungsethik-konsistent*
:: ‹((*'person*, *'world*) *handlungsabsicht* ⇒ *bool*) ⇒ (*'world handlung* ⇒ *bool*) ⇒ *bool*› **where**
‹*gesinnungsethik-verantwortungsethik-konsistent gesinnungsethik verantwortungsethik* ≡
∀ *handlungsabsicht*.
gesinnungsethik handlungsabsicht ⟷
(∀ *person welt*. *verantwortungsethik* (*handeln person welt handlungsabsicht*))›

Ich habe kein Beispiel für eine Gesinnungsethik und eine Verantwortungsethik, die tatsächlich konsistent sind.

4 Kant's Kategorischer Imperativ



„Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.“

https://de.wikipedia.org/wiki/Kategorischer_Imperativ

Meine persönliche, etwas utilitaristische, Interpretation.

5 Beispiel Person

Eine Beispielbevölkerung.

datatype *person* = *Alice* | *Bob* | *Carol* | *Eve*

Unsere Bevölkerung ist sehr endlich:

lemma *UNIV-person*: ‹*UNIV* = {*Alice*, *Bob*, *Carol*, *Eve*}›

Wir werden unterscheiden:

- *'person*: generischer Typ, erlaubt es jedes Modell einer Person und Bevölkerung zu haben.
- *person*: Unser minimaler Beispielstyp, bestehend aus *Alice*, *Bob*, ...

6 Maxime

Nach https://de.wikipedia.org/wiki/Maxime ist eine Maxime ein persönlicher Grundsatz des Wollens und Handelns. Nach Kant ist eine Maxime ein "subjektives Prinzip des Wollens".

Modell einer *Maxime*: Eine Maxime in diesem Modell beschreibt ob eine Handlung in einer gegebenen Welt gut ist.

Faktisch ist eine Maxime

- $'person$: die handelnde Person, i.e., *ich*.
- $'world\ handlung$: die zu betrachtende Handlung.
- *bool*: Das Ergebnis der Betrachtung. *True* = Gut; *False* = Schlecht.

Wir brauchen sowohl die $'world\ handlung$ als auch die $'person$ aus deren Sicht die Maxime definiert ist, da es einen großen Unterschied machen kann ob ich selber handel, ob ich Betroffener einer fremden Handlung bin, oder nur Außenstehender.

datatype $('person, 'world)\ maxime = Maxime\ \langle 'person \Rightarrow 'world\ handlung \Rightarrow bool\rangle$

Auswertung einer Maxime:

fun $okay :: \langle('person, 'world)\ maxime \Rightarrow 'person \Rightarrow 'world\ handlung \Rightarrow bool\rangle$ **where**
$\langle okay\ (Maxime\ m)\ p\ h = m\ p\ h\rangle$

Beispiel

definition $maxime\text{-}mir\text{-}ist\text{-}alles\text{-}recht :: \langle('person, 'world)\ maxime\rangle$ **where**
$\langle maxime\text{-}mir\text{-}ist\text{-}alles\text{-}recht \equiv Maxime\ (\lambda\text{- -}.\ True)\rangle$

6.1 Maxime in Sinne Kants?

Kants kategorischer Imperativ ist eine deontologische Ethik, d.h., "Es wird eben nicht bewertet, was die Handlung bewirkt, sondern wie die Absicht beschaffen ist." https://de.wikipedia.org/wiki/Kategorischer_Imperativ.

Wenn wir Kants kategorischen Imperativ bauen wollen, dürfen wir also nicht die Folgen einer Handlung betrachten, sondern nur die Absicht dahinter. Doch unsere *Maxime* betrachtet eine $'world\ handlung$, also eine konkrete Handlung, die nur durch ihre Folgen gegeben ist. Die Maxime betrachtet keine Handlungsabsicht $('person, 'world)\ handlungsabsicht$.

Dies mag nun als Fehler in unserem Modell verstanden werden. Doch irgendwo müssen wir praktisch werden. Nur von Handlungsabsichten zu reden, ohne dass die beabsichtigten Folgen betrachtet werden ist mir einfach zu abstrakt und nicht greifbar.

Kants kategorischer Imperativ und die Goldene Regel grundverschieden: https://web.archive.org/web/20220123174117/https://www.goethegymnasium-hildesheim.de/index.php/faecher/faecher/gesellschaftswissenschaften/philosophie Dennoch, betrachten wir den kategorischen Imperativ als eine Verallgemeinerung der goldenen Regel.

6.2 Die Goldene Regel

Die Goldene Regel nach https://de.wikipedia.org/wiki/Goldene_Regel sagt:

„Behandle andere so, wie du von ihnen behandelt werden willst.“

„Was du nicht willst, dass man dir tu, das füg auch keinem andern zu.“

So wie wir behandelt werden wollen ist modelliert durch eine $('person, 'world)\ maxime$.

Die goldene Regel testet ob eine Handlung, bzw. Handlungsabsicht moralisch ist. Um eine Handlung gegen eine Maxime zu testen fragen wir uns:

- Was wenn jeder so handeln würde?
- Was wenn jeder nach dieser Maxime handeln würde?

Beispielsweise mag "stehlen" und "bestohlen werden" die gleiche Handlung sein, jedoch wird sie von Täter und Opfer grundverschieden wahrgenommen.

definition *bevoelkerung* :: ‹*'person set*› **where** ‹*bevoelkerung* ≡ *UNIV*›
definition *wenn-jeder-so-handelt*
  :: ‹*'world* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ (*'world handlung*) *set*›
 where
  ‹*wenn-jeder-so-handelt welt handlungsabsicht* ≡
   (λ*handelnde-person*. *handeln handelnde-person welt handlungsabsicht*) ‘ *bevoelkerung*›
fun *was-wenn-jeder-so-handelt-aus-sicht-von*
  :: ‹*'world* ⇒ (*'person*, *'world*) *maxime* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ *'person* ⇒ *bool*›
 where
  ‹*was-wenn-jeder-so-handelt-aus-sicht-von welt m handlungsabsicht betroffene-person* =
    (∀ *h* ∈ *wenn-jeder-so-handelt welt handlungsabsicht*. *okay m betroffene-person h*)›

Für eine gegebene Welt und eine gegebene Maxime nennen wir eine Handlungsabsicht genau dann moralisch, wenn die Handlung auch die eigene Maxime erfüllt, wenn die Handlung von anderen durchgeführt würde.

definition *moralisch* ::
 ‹*'world* ⇒ (*'person*, *'world*) *maxime* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ *bool*› **where**
‹*moralisch welt handlungsabsicht maxime* ≡
 ∀*p* ∈ *bevoelkerung*. *was-wenn-jeder-so-handelt-aus-sicht-von welt handlungsabsicht maxime p*›

Faktisch bedeutet diese Definition, wir bilden das Kreuzprodukt Bevölkerung x Bevölkerung, wobei jeder einmal als handelnde Person auftritt und einmal als betroffene Person.

lemma *moralisch-unfold*:
 ‹*moralisch welt* (*Maxime m*) *handlungsabsicht* ⟷
    (∀*p1*∈*bevoelkerung*. ∀*p2*∈*bevoelkerung*. *m p1* (*handeln p2 welt handlungsabsicht*))›
lemma ‹*moralisch welt* (*Maxime m*) *handlungsabsicht* ⟷
    (∀(*p1*,*p2*)∈*bevoelkerung*×*bevoelkerung*. *m p1* (*handeln p2 welt handlungsabsicht*))›

lemma *moralisch-simp*:
 ‹*moralisch welt m handlungsabsicht* ⟷
    (∀*p1 p2*. *okay m p1* (*handeln p2 welt handlungsabsicht*))›

Wir können die goldene Regel auch umformulieren, nicht als Imperativ, sondern als Beobachtung eines Wunschzustandes: Wenn eine Handlung für eine Person okay ist, dann muss sie auch Okay sein, wenn jemand anderes diese Handlung ausführt.

Formal: $m\ ich\ (handeln\ ich\ welt\ handlungsabsicht) \Longrightarrow \forall p2.\ m\ ich\ (handeln\ p2\ welt\ handlungsabsicht)$

Genau dies können wir aus unserer Definition von *moralisch* ableiten:

lemma *goldene-regel*:
 ‹*moralisch welt m handlungsabsicht* ⟹
  okay m ich (*handeln ich welt handlungsabsicht*) ⟹

$\forall p2.$ *okay m ich* (*handeln p2 welt handlungsabsicht*)›

Für das obige lemma brauchen wir die Annahme *m ich* (*handeln ich welt handlungsabsicht*) gar nicht.

Wenn für eine gegebene *Maxime m* eine Handlungsabsicht moralisch ist, dann ist es auch okay, wenn ich von der Handlungsabsicht betroffen bin, egal wer sie ausführt.

corollary
‹*moralisch welt m handlungsabsicht* $\Longrightarrow$
$\forall p2.$ *okay m ich* (*handeln p2 welt handlungsabsicht*)›

Die umgekehrte Richtung gilt nicht, weil diese Formulierung nur die Handlungen betrachtet, die okay sind.

Hier schlägt das Programmiererherz höher: Wenn *'person* aufzählbar ist haben wir ausführbaren Code: *moralisch* = *moralisch-exhaust enum-class.enum* wobei *moralisch-exhaust* implementiert ist als *moralisch-exhaust bevoelk welt maxime handlungsabsicht* $\equiv$ *case maxime of Maxime m* $\Rightarrow$ *list-all* ($\lambda(p, x)$. *m p* (*handeln x welt handlungsabsicht*)) (*List.product bevoelk bevoelk*).

6.3 Maximen Debugging

Der folgende Datentyp modelliert ein Beispiel in welcher Konstellation eine gegebene Maxime verletzt ist:

datatype *'person opfer* = *Opfer* ‹*'person*›
datatype *'person taeter* = *Taeter* ‹*'person*›
datatype (*'person*, *'world*) *verletzte-maxime* =
VerletzteMaxime
‹*'person opfer*› — verletzt für; das Opfer
‹*'person taeter*› — handelnde Person; der Täter
‹*'world handlung*› — Die verletzende Handlung

Die folgende Funktion liefert alle Gegebenheiten welche eine Maxime verletzen:

fun *debug-maxime*
:: ‹(*'world* $\Rightarrow$ *'printable-world*) $\Rightarrow$ *'world* $\Rightarrow$
(*'person*, *'world*) *maxime* $\Rightarrow$ (*'person*, *'world*) *handlungsabsicht*
$\Rightarrow$ ((*'person*, *'printable-world*) *verletzte-maxime*) *set*›
where
‹*debug-maxime print-world welt m handlungsabsicht* =
{*VerletzteMaxime*
(*Opfer p1*) (*Taeter p2*)
(*map-handlung print-world* (*handeln p2 welt handlungsabsicht*)) | *p1 p2*.
$\neg$*okay m p1* (*handeln p2 welt handlungsabsicht*)}›

Es gibt genau dann keine Beispiele für Verletzungen, wenn die Maxime erfüllt ist:

lemma ‹*debug-maxime print-world welt maxime handlungsabsicht* = {}
$\longleftrightarrow$ *moralisch welt maxime handlungsabsicht*›

6.4 Beispiel

Beispiel: Die Welt sei nur eine Zahl und die zu betrachtende Handlungsabsicht sei, dass wir diese Zahl erhöhen. Die Mir-ist-alles-Recht Maxime ist hier erfüllt:

```
lemma ‹moralisch
        (42::nat)
         maxime-mir-ist-alles-recht
        (Handlungsabsicht (λ(person::person) welt. welt + 1))›
```

Beispiel: Die Welt ist modelliert als eine Abbildung von Person auf Besitz. Die Maxime sagt, dass Leute immer mehr oder gleich viel wollen, aber nie etwas verlieren wollen. In einer Welt in der keiner etwas hat, erfüllt die Handlung jemanden 3 zu geben die Maxime.

```
lemma ‹moralisch
        [Alice ↦ (0::nat), Bob ↦ 0, Carol ↦ 0, Eve ↦ 0]
        (Maxime (λperson handlung.
            (the ((vorher handlung) person)) ≤ (the ((nachher handlung) person))))
        (Handlungsabsicht (λperson welt. welt(person ↦ 3)))›
lemma ‹debug-maxime show-map
        [Alice ↦ (0::nat), Bob ↦ 0, Carol ↦ 0, Eve ↦ 0]
        (Maxime (λperson handlung.
            (the ((vorher handlung) person)) ≤ (the ((nachher handlung) person))))
        (Handlungsabsicht (λperson welt. welt(person ↦ 3)))
  = {}›
```

Wenn nun *Bob* allerdings bereits 4 hat, würde die obige Handlung ein Verlust für ihn bedeuten und die Maxime ist nicht erfüllt.

```
lemma ‹¬ moralisch
        [Alice ↦ (0::nat), Bob ↦ 4, Carol ↦ 0, Eve ↦ 0]
        (Maxime (λperson handlung.
            (the ((vorher handlung) person)) ≤ (the ((nachher handlung) person))))
        (Handlungsabsicht (λperson welt. welt(person ↦ 3)))›
lemma ‹debug-maxime show-map
        [Alice ↦ (0::nat), Bob ↦ 4, Carol ↦ 0, Eve ↦ 0]
        (Maxime (λperson handlung.
            (the ((vorher handlung) person)) ≤ (the ((nachher handlung) person))))
        (Handlungsabsicht (λperson welt. welt(person ↦ 3)))
  = {VerletzteMaxime (Opfer Bob) (Taeter Bob)
    (Handlung [(Alice, 0), (Bob, 4), (Carol, 0), (Eve, 0)]
          [(Alice, 0), (Bob, 3), (Carol, 0), (Eve, 0)])}›
```

7 Schleier des Nichtwissens

Rawls' Schleier des Nichtwissens https://de.wikipedia.org/wiki/Schleier_des_Nichtwissens ist ein fiktives Modell, » über die zukünftige Gesellschaftsordnung entscheiden können, aber selbst nicht wissen, an welcher Stelle dieser Ordnung sie sich später befinden werden, also unter einem „Schleier des Nichtwissens“ stehen.« Quote wikipedia

Wir bedienen uns bei der Idee dieses Modells um gültige Handlungsabsichten und Maximen zu definieren. Handlungsabsichten und Maximen sind nur gültig, wenn darin keine Personen hardgecoded werden.

Beispielsweise ist folgende Handlungsabsicht ungültig: λ*ich welt. if ich* = *Alice then Do-A welt else Do-B welt*

Handlungsabsichten und Maximen müssen immer generisch geschrieben werden, so dass die handelnden und betroffenen Personen niemals anhand ihres Namens ausgewählt werden.

unser Modell von Handlungsabsichten und Maximen stellt beispielsweise die handelnde Person als Parameter bereit. Folgendes ist also eine gültige Handlung: λ*ich welt. ModifiziereWelt welt ich*

Auch ist es erlaubt, Personen in einer Handlungsabsicht oder Maxime nur anhand ihrer Eigenschaften in der Welt auszuwählen. Folgendes wäre eine wohlgeformte Handlung, wenn auch eine moralisch fragwürdige: λ*ich welt. enteignen ‘ {opfer. besitz ich* < *besitz opfer}*

Um diese Idee von wohlgeformten Handlungsabsichten und Maximen zu formalisieren bedienen wir uns der Idee des Schleiers des Nichtwissens. Wir sagen, dass Handlungsabsichten wohlgeformt sind, wenn die Handlungsabsicht gleich bleibt, wenn man sowohl die handelnde Person austauscht, als auch alle weltlichen Eigenschaften dieser Person. Anders ausgedrückt: Wohlgeformte Handlungsabsichten und Maximen sind solche, bei denen bei der Definition noch nicht feststeht, auf we sie später zutreffen.

Für jede Welt muss eine Welt-Personen Swap (wps) Funktion bereit gestellt werden, die alle Weltlichen Eigenschaften von 2 Personen vertauscht:

type-synonym (*'person*, *'world*) *wp-swap* = ‹*'person* ⇒ *'person* ⇒ *'world* ⇒ *'world*›

$$\begin{tikzcd} welt \arrow[r, "handeln\ p1"] \arrow[d, "welt\text{-}personen\text{-}swap\ p1\ p2"] & welt' \\ alternativ\text{-}welt \arrow[r, "handeln\ p2"] & alternativ\text{-}welt' \arrow[u, "welt\text{-}personen\text{-}swap\ p1\ p2"'] \end{tikzcd}$$

7.1 Wohlgeformte Handlungsabsicht

definition *wohlgeformte-handlungsabsicht*
  :: ‹(*'person*, *'world*) *wp-swap* ⇒ *'world* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ *bool*›
where
  ‹*wohlgeformte-handlungsabsicht wps welt h* ≡
    ∀ *p1 p2. handeln p1 welt h* =
        map-handlung (*wps p2 p1*) (*handeln p2* (*wps p1 p2 welt*) *h*)›

Folgende Equivalenz erklärt die Definition vermutlich besser:

lemma *wohlgeformte-handlungsabsicht-simp*:
  ‹*wohlgeformte-handlungsabsicht wps welt h* ⟷
    (∀ *p1 p2. wps p2 p1* (*wps p1 p2 welt*) = *welt*) ∧
    (∀ *p1 p2. handeln p1 welt h* =
        Handlung welt

(*wps p2 p1* (*nachher* (*handeln p2* (*wps p1 p2 welt*) *h*))))›

definition *wohlgeformte-handlungsabsicht-gegenbeispiel*
:: ‹('*person*, '*world*) *wp-swap* ⇒ '*world* ⇒ ('*person*, '*world*) *handlungsabsicht* ⇒ '*person* ⇒ '*person* ⇒ *bool*›
where
‹*wohlgeformte-handlungsabsicht-gegenbeispiel wps welt h taeter opfer* ≡
handeln taeter welt h ≠
map-handlung (*wps opfer taeter*) (*handeln opfer* (*wps taeter opfer welt*) *h*)›

lemma ‹*wohlgeformte-handlungsabsicht-gegenbeispiel wps welt h p1 p2* ⟹
¬*wohlgeformte-handlungsabsicht wps welt h*›

lemma *wohlgeformte-handlungsabsicht-imp-swpaid*:
‹*wohlgeformte-handlungsabsicht wps welt h* ⟹
wps p1 p2 (*wps p2 p1 welt*) = *welt*›

Nach der gleichen Argumentation müssen Maxime und Handlungsabsicht so generisch sein, dass sie in allen Welten zum gleichen Ergebnis kommen.

definition *maxime-und-handlungsabsicht-generalisieren*
:: ‹('*person*, '*world*) *wp-swap* ⇒ '*world* ⇒
('*person*, '*world*) *maxime* ⇒ ('*person*, '*world*) *handlungsabsicht* ⇒ '*person* ⇒ *bool*›
where
‹*maxime-und-handlungsabsicht-generalisieren wps welt m h p* =
(∀ *p1 p2*. (¬*ist-noop* (*handeln p welt h*) ∧ ¬*ist-noop* (*handeln p* (*wps p1 p2 welt*) *h*))
⟶ *okay m p* (*handeln p welt h*) ⟷ *okay m p* (*handeln p* (*wps p1 p2 welt*) *h*))›

Für eine gegebene Maxime schließt die Forderung *maxime-und-handlungsabsicht-generalisieren* leider einige Handlungen aus. Beispiel: In einer Welt besitzt *Alice* 2 und *Eve* hat 1 Schulden. Die Maxime ist, dass Individuen gerne keinen Besitz verlieren. Die Handlung sei ein globaler reset, bei dem jeden ein Besitz von 0 zugeordnet wird. Leider generalisiert diese Handlung nicht, da *Eve* die Handlung gut findet, *Alice* allerdings nicht.

lemma
‹¬ *maxime-und-handlungsabsicht-generalisieren*
swap
((λ*x*. *0*)(*Alice* := (*2*::*int*), *Eve* := − *1*))
(*Maxime* (λ*ich h*. (*vorher h*) *ich* ≤ (*nachher h*) *ich*))
(*Handlungsabsicht* (λ*ich w*. (λ-. *0*)))
Eve›

Die Maxime und ('*person*, '*world*) *wp-swap* müssen einige Eigenschaften erfüllen. Wir kürzen das ab mit *wpsm*: Welt Person Swap Maxime.

Die Person für die Maxime ausgewertet wird und swappen der Personen in der Welt muss equivalent sein:

definition *wpsm-kommutiert*
  :: ‹('person, 'world) maxime ⇒ ('person, 'world) wp-swap ⇒ 'world ⇒ bool›
where
  ‹*wpsm-kommutiert m wps welt* ≡
∀ *p1 p2 h*.
  okay m p2 (*Handlung* (*wps p1 p2 welt*) (*h p1* (*wps p1 p2 welt*)))
  ⟷
  okay m p1 (*Handlung welt* (*wps p1 p2* (*h p1* (*wps p2 p1 welt*))))›

lemma *wpsm-kommutiert-simp*: ‹*wpsm-kommutiert m wps welt* =
(∀ *p1 p2 h*.
  okay m p2 (*handeln p1* (*wps p1 p2 welt*) (*Handlungsabsicht h*))
  ⟷
  okay m p1 (*handeln p1 welt* (*Handlungsabsicht* (λ*p w*. *wps p1 p2* (*h p* (*wps p2 p1 w*)))))
)›

Wenn sowohl *wohlgeformte-handlungsabsicht* als auch *wpsm-kommutiert*, dann erhalten wir ein sehr intuitives Ergebnis, welches besagt, dass ich handelnde Person und Person für die die Maxime gelten soll vertauschen kann.

lemma *wfh-wpsm-kommutiert-simp*:
  ‹*wohlgeformte-handlungsabsicht wps welt ha* ⟹
  wpsm-kommutiert m wps welt ⟹
    okay m p2 (*handeln p1* (*wps p1 p2 welt*) *ha*)
    ⟷
    okay m p1 (*handeln p2 welt ha*)›

Die Rückrichtung gilt auch, aber da wir das für alle Handlungsabsichten in der Annahme brauchen, ist das eher weniger hilfreich.

lemma *wfh-kommutiert-wpsm*:
  ‹∀ *ha*. *wohlgeformte-handlungsabsicht wps welt ha* ∧
      (∀ *p1 p2*. *okay m p2* (*handeln p1* (*wps p1 p2 welt*) *ha*)
          ⟷
          okay m p1 (*handeln p2 welt ha*)) ⟹
wpsm-kommutiert m wps welt›

7.2 Wohlgeformte Maxime

definition *wohlgeformte-maxime-auf*
  :: ‹'world handlung ⇒ ('person, 'world) wp-swap ⇒ ('person, 'world) maxime ⇒ bool›
where
  ‹*wohlgeformte-maxime-auf h wps m* ≡
    ∀ *p1 p2*. *okay m p1 h* ⟷ *okay m p2* (*map-handlung* (*wps p1 p2*) *h*)›

definition *wohlgeformte-maxime*
  :: ‹('person, 'world) wp-swap ⇒ ('person, 'world) maxime ⇒ bool›
where

$\langle$*wohlgeformte-maxime wps m* $\equiv$
$\forall$ *h. wohlgeformte-maxime-auf h wps m*$\rangle$

Beispiel:

lemma $\langle$*wohlgeformte-maxime swap* (*Maxime* (λ*ich h.* (*vorher h*) *ich* $\leq$ (*nachher h*) *ich*))$\rangle$

8 Kategorischer Imperativ

Wir haben mit der goldenen Regel bereits definiert, wann für eine gegebene Welt und eine gegebene maxime, eine Handlungsabsicht moralisch ist:

- *moralisch*::*'world* $\Rightarrow$ (*'person*, *'world*) *maxime* $\Rightarrow$ (*'person*, *'world*) *handlungsabsicht* $\Rightarrow$ *bool*

Effektiv testet die goldene Regel eine Handlungsabsicht.
Nach meinem Verständnis generalisiert Kant mit dem Kategorischen Imperativ diese Regel, indem die Maxime nicht mehr als gegeben angenommen wird, sondern die Maxime selbst getestet wird. Sei die Welt weiterhin gegeben, dass müsste der kategorische Imperativ folgende Typsignatur haben:

- *'world* $\Rightarrow$ (*'person*, *'world*) *maxime* $\Rightarrow$ *bool*

Eine Implementierung muss dann über alle möglichen Handlungsabsichten allquantifizieren.
Ich behaupte, der kategorischer Imperativ lässt sich wie folgt umformulieren:

- Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.
- Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie jeder befolgt, im Sinne der goldenen Regel.
- Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie (Handlung+Maxime) moralisch ist.
- Wenn es jemanden gibt der nach einer Maxime handeln will, dann muss diese Handlung nach der Maxime moralisch sein.
- Für jede Handlungsabsicht muss gelten: Wenn jemand in jeder Welt nach der Handlungsabsicht handeln würde, dann muss diese Handlung moralisch sein.

Daraus ergibt sich diese Formalisierung:
Für eine bestimmte Handlungsabsicht: Wenn es eine Person gibt für die diese Handlungsabsicht moralisch ist, dann muss diese Handlungsabsicht auch für alle moralisch (im Sinne der goldenen Regel) sein.

definition *kategorischer-imperativ-auf*

:: ‹('*person*, '*world*) *handlungsabsicht* ⇒ '*world* ⇒ ('*person*, '*world*) *maxime* ⇒ *bool*›
where
‹*kategorischer-imperativ-auf h welt m* ≡
(∃ *ich*. ¬*ist-noop* (*handeln ich welt h*) ∧ *okay m ich* (*handeln ich welt h*)) ⟶ *moralisch welt m h*›

Für alle möglichen (wohlgeformten) Handlungsabsichten muss dies nun gelten:

definition *kategorischer-imperativ*
:: ‹('*person*, '*world*) *wp-swap* ⇒ '*world* ⇒ ('*person*, '*world*) *maxime* ⇒ *bool*›
where
‹*kategorischer-imperativ wps welt m* ≡
∀ *h*. *wohlgeformte-handlungsabsicht wps welt h* ⟶
kategorischer-imperativ-auf h welt m›

Wir führen die interne Hilfsdefinition *kategorischer-imperativ-auf* ein um den kategorischen Imperativ nur für eine Teilmenge aller Handlungen besser diskutieren zu können.

TODO: Leider fehlen mir Beispiele von Maximen welche den kategorischen Imperativ uneingeschränkt auf allen Handlungsabsichten erfüllen.

Diese ¬ *ist-noop* (*handeln ich welt h*) gefällt mir gar nicht. Wir brauchen es aber, damit die Beispiele funktionieren. Das ist nötig, um pathologische Grenzfälle auszuschließen. Beispielsweise ist von-sich-selbst stehlen eine no-op. No-ops sind normalerweise nicht böse. Stehlen ist schon böse. Dieser Grenzfall in dem Stehlen zur no-op wird versteckt also den Charakter der Handlungsabsicht und muss daher ausgeschlossen werden. Ganz glücklich bin ich mit der Rechtfertigung aber nicht. Eventuell wäre es schöner, Handlungen partiell zu machen, also dass Handlungsabsichten auch mal *None* zurückgeben dürfen. Das könnte einiges rechtfertigen. Beispielsweise ist Stehlen: jemand anderen etwas wegnehmen. Nicht von sich selbst. Allerdings machen partielle Handlungen alles komplizierter.

In der Definition is *wohlgeformte-handlungsabsicht* ein technisch notwendiges Implementierungsdetail um nicht-wohlgeformte Handlungen auszuschließen.

Minimal andere Formulierung:

lemma
‹*kategorischer-imperativ wps welt m* ⟷
(∀ *h*.
(∃ *p*.
wohlgeformte-handlungsabsicht wps welt h ∧
¬*ist-noop* (*handeln p welt h*) ∧
okay m p (*handeln p welt h*))
⟶ *moralisch welt m h*)›

Der Existenzquantor lässt sich auch in einen Allquantor umschreiben:

lemma
‹*kategorischer-imperativ wps welt m* ⟷
(∀ *h ich*.
wohlgeformte-handlungsabsicht wps welt h ∧ ¬*ist-noop* (*handeln ich welt h*) ∧
okay m ich (*handeln ich welt h*) ⟶ *moralisch welt m h*)›

Vergleich zu *moralisch*. Wenn eine Handlung moralisch ist, dann impliziert diese Handlung die Kernforderung des *kategorischer-imperativ*. Wenn die Handlungsabsicht für mich okaz ist, ist sie auch für alle anderen okay.

lemma ‹*moralisch welt m ha* ⟹
    kategorischer-imperativ-auf ha welt m›

Die andere Richtung gilt nicht, z.B. ist die Maxime die immer False zurückgibt ein Gegenbeispiel.

lemma ‹*m* = *Maxime* (λ- -. *False*) ⟹
    kategorischer-imperativ-auf ha welt m ⟶ *moralisch welt m ha*
      ⟹ *False*›

Für jede Handlungsabsicht: wenn ich so handeln würde muss es auch okay sein, wenn zwei beliebige Personen so handeln, wobei einer Täter und einer Opfer ist.

lemma *kategorischer-imperativ-simp*:
  ‹*kategorischer-imperativ wps welt m* ⟷
   (∀ *ha p1 p2 ich*.
     wohlgeformte-handlungsabsicht wps welt ha ∧ ¬*ist-noop* (*handeln ich welt ha*) ∧
     okay m ich (*handeln ich welt ha*)
     ⟶ *okay m p1* (*handeln p2 welt ha*))›

Introduction rules

lemma *kategorischer-imperativI*:
  ‹(⋀*ha ich p1 p2*. *wohlgeformte-handlungsabsicht wps welt ha* ⟹
        ¬*ist-noop* (*handeln ich welt ha*) ⟹
        okay m ich (*handeln ich welt ha*) ⟹ *okay m p1* (*handeln p2 welt ha*)
)
     ⟹ *kategorischer-imperativ wps welt m*›

lemma *kategorischer-imperativ-aufI*:
  ‹(⋀*ich p1 p2*. ¬*ist-noop* (*handeln ich welt ha*)
     ⟹ *okay m ich* (*handeln ich welt ha*) ⟹ *okay m p1* (*handeln p2 welt ha*))
     ⟹ *kategorischer-imperativ-auf ha welt m*›

8.1 Triviale Maximen die den Kategorischen Imperativ immer Erfüllen

Die Maxime die keine Handlung erlaubt (weil immer False) erfüllt den kategorischen Imperativ:

lemma ‹*kategorischer-imperativ wps welt* (*Maxime* (λ*ich h*. *False*))›

Allerdings kann mit so einer Maxime nie etwas moralisch sein.

lemma ‹¬ *moralisch welt* (*Maxime* (λ*ich h*. *False*)) *h*›

Die Maxime die jede Handlung erlaubt (weil immer True) erfüllt den kategorischen Imperativ:

lemma ‹*kategorischer-imperativ wps welt* (*Maxime* (λ*ich h. True*))›

Allerdings ist mot so einer Maxime alles moralisch.

lemma ‹*moralisch welt* (*Maxime* (λ*ich h. True*)) *h*›

8.2 Zusammenhang Goldene Regel

Mit der goldenen Regel konnten wir wie folgt moralische Entscheidungen treffen: $[\![$*moralisch welt m handlungsabsicht*; *okay m ich* (*handeln ich welt handlungsabsicht*)$]\!] \Longrightarrow \forall$ *p2*. *okay m ich* (*handeln p2 welt handlungsabsicht*)

In Worten: Wenn eine Handlungsabsicht moralisch ist (nach goldener Regel) und es okay ist für mich diese Handlung auszuführen, denn ist es auch für mich okay, wenn jeder andere diese Handlung mit mir als Opfer ausführt.

Der kategorische Imperativ liftet dies eine Abstraktionsebene:

lemma ‹*kategorischer-imperativ wps welt m* $\Longrightarrow$
wohlgeformte-handlungsabsicht wps welt ha $\Longrightarrow$
$\neg$*ist-noop* (*handeln ich welt ha*) $\Longrightarrow$
okay m ich (*handeln ich welt ha*) $\Longrightarrow$ *moralisch welt m ha*›

In Worten: Wenn eine Maxime den kategorischen Imperativ erfüllt und es für eine beliebige (wohlgeformte) Handlung auszuführen für mich okay ist diese auszuführen, dann ist diese Handlung moralisch..

Für Beispiele wird es einfacher zu zeigen, dass eine Maxime nicht den kategorischen Imperativ erfüllt, wenn wir direkt ein Beispiel angeben.

definition ‹*kategorischer-imperativ-gegenbeispiel wps welt m ha ich p1 p2* $\equiv$
wohlgeformte-handlungsabsicht wps welt ha $\wedge$
$\neg$*ist-noop* (*handeln ich welt ha*) $\wedge$ *okay m ich* (*handeln ich welt ha*) $\wedge$
$\neg$ *okay m p1* (*handeln p2 welt ha*)›

lemma ‹*kategorischer-imperativ-gegenbeispiel wps welt m ha ich p1 p2* $\Longrightarrow$
$\neg$ *kategorischer-imperativ wps welt m*›

8.3 Maximen die den Kategorischen Imperativ immer Erfüllen

Wenn eine Maxime jede Handlungsabsicht als moralisch bewertet, erfüllt diese Maxime den kategorischen Imperativ. Da diese Maxime jede Handlung erlaubt, ist es dennoch eine wohl ungeeignete Maxime.

lemma ‹$\forall$ *ha*. *moralisch welt maxime ha* $\Longrightarrow$ *kategorischer-imperativ wps welt maxime*›

Eine Maxime die das ich und die Handlung ignoriert erfüllt den kategorischen Imperativ.

lemma *blinde-maxime-katimp*:
‹*kategorischer-imperativ wps welt* (*Maxime* (λ*ich h. m*))›

Eine Maxime welche das *ich* ignoriert, also nur die Handlung global betrachtet, erfüllt den kategorischen Imperativ.

theorem *globale-maxime-katimp*:
fixes *P* :: ‹*'world handlung* ⇒ *bool*›
assumes *mhg*: ‹∀ *p. maxime-und-handlungsabsicht-generalisieren wps welt* (*Maxime* (λ*ich*::*'person. P*)) *ha p*›
and *maxime-erlaubt-untaetigkeit*: ‹∀ *p. ist-noop* (*handeln p welt ha*) ⟶ *okay* (*Maxime* (λ*ich*::*'person. P*)) *p* (*handeln p welt ha*)›
and *kom*: ‹*wpsm-kommutiert* (*Maxime* (λ*ich*::*'person. P*)) *wps welt*›
and *wps-sym*:
‹∀ *p1 p2 welt. wps p1 p2 welt* = *wps p2 p1 welt*›
and *wps-id*:
‹∀ *p1 p2 welt. wps p1 p2* (*wps p1 p2 welt*) = *welt*›
and *wfh*: ‹*wohlgeformte-handlungsabsicht wps welt ha*›
shows ‹*kategorischer-imperativ-auf ha welt* (*Maxime* (λ*ich*::*'person. P*))›

9 Experimental: Beispiel

value‹[(*x*,*y*). *x* ← *xs*, *y* ← *ys*, *x* ≠ *y*]›

definition *alle-moeglichen-handlungen*
:: ‹*'world* ⇒ (*'person*::*enum*, *'world*) *handlungsabsicht list* ⇒ *'world handlung list*›
where
‹*alle-moeglichen-handlungen welt has* ≡ [*handeln p welt ha. ha* ← *has*, *p* ← (*Enum.enum*::*'person list*)]›

lemma *set-alle-moeglichen-handlungen*:
‹*set* (*alle-moeglichen-handlungen welt has*) = {*handeln p welt ha* | *ha p. ha*∈*set has*}›

record (*'person*, *'world*) *beipiel* =
bsp-welt :: ‹*'world*›
bsp-erfuellte-maxime :: ‹(*'person*, *'world*) *maxime option*›
bsp-erlaubte-handlungen :: ‹(*'person*, *'world*) *handlungsabsicht list*›
bsp-verbotene-handlungen :: ‹(*'person*, *'world*) *handlungsabsicht list*›

definition *erzeuge-beispiel*
:: ‹(*'person*::*enum*, *'world*) *wp-swap* ⇒ *'world* ⇒
(*'person*, *'world*) *handlungsabsicht list* ⇒ (*'person*, *'world*) *maxime*
⇒ (*'person*, *'world*) *beipiel option*›
where

‹*erzeuge-beispiel wps welt has m* ≡
  if ($\exists$ *h*∈*set* (*alle-moeglichen-handlungen welt has*). ¬*wohlgeformte-maxime-auf h wps m*)
    ∨ ($\exists$ *ha*∈*set has*. ¬ *wohlgeformte-handlungsabsicht wps welt ha*)
  then None
  else Some
  ⦇ *bsp-welt* = *welt*,
    bsp-erfuellte-maxime = *if* ∀ *ha*∈*set has*. *kategorischer-imperativ-auf ha welt m then Some m else None*,
    bsp-erlaubte-handlungen = [*ha*←*has*. *moralisch welt m ha*],
    bsp-verbotene-handlungen = [*ha*←*has*. ¬ *moralisch welt m ha*]
  ⦈›

erzeuge-beispiel erzeugt nur ein Beiespiel wenn alles wohlgeformt ist.

lemma *erzeuge-beispiel wps welt has m* = *Some bsp* ⟹
  (∀ *ha* ∈ *set has*. *wohlgeformte-handlungsabsicht wps welt ha*) ∧
  (∀ *h* ∈ *set* (*alle-moeglichen-handlungen welt has*). *wohlgeformte-maxime-auf h wps m*)

lemma‹*erzeuge-beispiel swap* (λ*p*::*person*. *0*::*int*) [*Handlungsabsicht* (λ*p w*. *w*)] (*Maxime* (λ*ich w*. *True*))
  =
  Some
  ⦇*bsp-welt* = (λ*p*::*person*. *0*::*int*),
   bsp-erfuellte-maxime = *Some* (*Maxime* (λ*ich w*. *True*)),
   bsp-erlaubte-handlungen = [*Handlungsabsicht* (λ*p w*. *w*)],
   bsp-verbotene-handlungen = []
  ⦈›

10 Utilitarismus

Wir betrachten hier primär einen einfachen Handlungsutilitarismus. Frei nach Jeremy Bentham. Sehr frei. Also sehr viel persönliche Auslegung.

Eine Handlung ist genau dann moralisch richtig, wenn sie den aggregierten Gesamtnutzen, d.h. die Summe des Wohlergehens aller Betroffenen, maximiert wird.

type-synonym *'world glueck-messen* = ‹*'world handlung* ⇒ *ereal*›

Wir messen Glück im Typen *ereal*, also reelle Zahlen mit ∞ und $-\infty$, so dass auch "den höchsten Preis zahlen" modelliert werden kann.

lemma ‹(λ*h*::*ereal handlung*. *case h of Handlung vor nach* ⇒ *nach* − *vor*) (*Handlung 3 5*) = *2*›
lemma ‹(λ*h*::*ereal handlung*. *case h of Handlung vor nach* ⇒ *nach* − *vor*) (*Handlung 3* ∞) = ∞›
lemma ‹(λ*h*::*ereal handlung*. *case h of Handlung vor nach* ⇒ *nach* − *vor*) (*Handlung 3* (−∞)) = −∞›

definition *moralisch-richtig* :: ‹*'world glueck-messen* ⇒ *'world handlung* ⇒ *bool*› **where**
  ‹*moralisch-richtig glueck-messen handlung* ≡ (*glueck-messen handlung*) ≥ *0*›

10.1 Goldene Regel und Utilitarismus im Einklang

In diese kleinen Intermezzo werden wir zeigen, wie sich die Gesinnungsethik der goldenen Regel in die Verantwortungsethik des Utilitarismus übersetzen lässt.

definition *goldene-regel-als-gesinnungsethik*
  :: ‹(*'person*, *'world*) *maxime* ⇒ (*'person*, *'world*) *handlungsabsicht* ⇒ *bool*›
where
  ‹*goldene-regel-als-gesinnungsethik maxime handlungsabsicht* ≡
    ∀ *welt*. *moralisch welt maxime handlungsabsicht*›

definition *utilitarismus-als-verantwortungsethik*
  :: ‹*'world glueck-messen* ⇒ *'world handlung* ⇒ *bool*›
where
  ‹*utilitarismus-als-verantwortungsethik glueck-messen handlung* ≡
    moralisch-richtig glueck-messen handlung›

Eine Maxime ist immer aus Sicht einer bestimmten Person definiert. Wir "neutralisieren" eine Maxime indem wir diese bestimmte Person entfernen und die Maxime so allgemeingültiger machen. Alle Personen müssen gleich behandelt werden Um die Maxime unabhängig von einer bestimmten Person zu machen, fordern wir einfach, dass die Maxime für aller Personen erfüllt sein muss.

fun *maximeNeutralisieren* :: ‹(*'person*, *'world*) *maxime* ⇒ (*'world handlung* ⇒ *bool*)› **where**
  ‹*maximeNeutralisieren* (*Maxime m*) = (λ*welt*. ∀ *p*::*'person*. *m p welt*)›

Nun übersetzen wir eine Maxime in die *'world glueck-messen* Funktion des Utilitarismus. Der Trick: eine verletzte Maxime wird als unendliches Leid übersetzt.

definition *maxime-als-nutzenkalkuel*
  :: ‹(*'person*, *'world*) *maxime* ⇒ *'world glueck-messen*›
where
  ‹*maxime-als-nutzenkalkuel maxime* ≡
    (λ*welt*. *case* (*maximeNeutralisieren maxime*) *welt*
        of True ⇒ *1*
        | *False* ⇒ − ∞)›

Für diese Übersetzung können wir beweisen, dass die Gesinnungsethik der goldenen Regel und die utilitaristische Verantwortungsethik konsistent sind:

theorem ‹*gesinnungsethik-verantwortungsethik-konsistent*
    (*goldene-regel-als-gesinnungsethik maxime*)
    (*utilitarismus-als-verantwortungsethik* (*maxime-als-nutzenkalkuel maxime*))›

Diese Konsistenz gilt nicht im allgemeinen, sondern nur wenn Glück gemessen wird mit Hilfe der *maxime-als-nutzenkalkuel* Funktion. Der Trick dabei ist nicht, dass wir einer verletzten Maxime $-\infty$ Nutzen zuordnen, sondern der Trick besteht in *maximeNeutralisieren*, welche nicht erlaubt Glück aufzuaddieren und mit Leid zu verrechnen, sondern dank des Allquantors dafür sorgt, dass auch nur das kleinste Leid dazu führt, dass sofort *False* zurückgegebn wird.

Aber wenn wir ordentlich aufsummieren, jedoch einer verletzten Maxime $-\infty$ Nutzen zuordnen und zusätzlich annehmen, dass die Bevölkerung endlich ist, dann funktioniert das auch:

```
fun maxime-als-summe-wohlergehen
  :: ‹('person, 'world) maxime ⇒ 'world glueck-messen›
where
  ‹maxime-als-summe-wohlergehen (Maxime m) =
    (λwelt. ∑p∈bevoelkerung. (case m p welt
                                of True ⇒ 1
                                | False ⇒ − ∞))›
theorem
  fixes maxime :: ‹('person, 'world) maxime›
  assumes ‹finite (bevoelkerung:: 'person set)›
  shows
    ‹gesinnungsethik-verantwortungsethik-konsistent
      (goldene-regel-als-gesinnungsethik maxime)
      (utilitarismus-als-verantwortungsethik (maxime-als-summe-wohlergehen maxime))›
```

11 Zahlenwelt Helper

Wir werden Beispiele betrachten, in denen wir Welten modellieren, in denen jeder Person eine Zahl zugewiesen wird: *person ⇒ int*. Diese Zahl kann zum Beispiel der Besitz oder Wohlstand einer Person sein, oder das Einkommen. Wobei Gesamtbesitz und Einkommen über einen kurzen Zeitraum recht unterschiedliche Sachen modellieren.
Hier sind einige Hilfsfunktionen um mit *person ⇒ int* allgemein zu arbeiten.

Default: Standardmäßig hat jede Person *0*:

```
definition DEFAULT :: ‹person ⇒ int› where
  ‹DEFAULT ≡ λp. 0›
```

Beispiel:

```
lemma ‹(DEFAULT(Alice:=8, Bob:=3, Eve:= 5)) Bob = 3›
```

Beispiel mit fancy Syntax:

```
lemma ‹🌍[Alice:=8, Bob:=3, Eve:= 5] Bob = 3›

lemma ‹show-fun 🌍[Alice := 4, Carol := 4] = [(Alice, 4), (Bob, 0), (Carol, 4), (Eve, 0)]›
lemma ‹show-num-fun 🌍[Alice := 4, Carol := 4] = [(Alice, 4), (Carol, 4)]›
```

```
abbreviation num-fun-add-syntax (- '(- += -')) where
  ‹f(p += n) ≡ (f(p := (f p) + n))›
```

abbreviation *num-fun-minus-syntax* (- $'$(- $-{=}$ -$'$)) **where**
‹$f(p\ -{=}\ n) \equiv (f(p := (f\ p) - n))$›

lemma ‹(🌍[*Alice*:=*8*, *Bob*:=*3*, *Eve*:= *5*])(*Bob* += *4*) *Bob* = *7*›
lemma ‹(🌍[*Alice*:=*8*, *Bob*:=*3*, *Eve*:= *5*])(*Bob* $-{=}$ *4*) *Bob* = $-$*1*›

lemma fixes *n*:: ‹*int*› **shows** ‹$f(p\ {+}{=}\ n)(p\ -{=}\ n) = f$›

Diskriminierungsfrei eine *'person* eindeutig anhand Ihres Besitzes auswählen:

definition *opfer-eindeutig-nach-besitz-auswaehlen*
:: ‹*int* ⇒ (*'person* ⇒ *int*) ⇒ *'person list* ⇒ *'person option*› **where**
‹*opfer-eindeutig-nach-besitz-auswaehlen b besitz ps* =
(*case filter* (λ*p*. *besitz p* = *b*) *ps*
of [*opfer*] ⇒ *Some opfer*
| - ⇒ *None*)›
definition *the-single-elem* :: ‹*'a set* ⇒ *'a option*› **where**
‹*the-single-elem s* ≡ *if card s* = *1 then Some* (*Set.the-elem s*) *else None*›

thm *is-singleton-the-elem*[*symmetric*]
lemma ‹*A* = {*the-elem A*} ⟷ *is-singleton A*›

lemma *opfer-nach-besitz-induct-step-set-simp*: ‹*besitz a* ≠ *opfer-nach-besitz* ⟹
{*p*. (*p* = *a* ∨ *p* ∈ *set ps*) ∧ *besitz p* = *opfer-nach-besitz*} =
{*p* ∈ *set ps*. *besitz p* = *opfer-nach-besitz*}›

lemma *opfer-eindeutig-nach-besitz-auswaehlen-the-single-elem*:
‹*distinct ps* ⟹
opfer-eindeutig-nach-besitz-auswaehlen opfer-nach-besitz besitz ps =
the-single-elem {*p* ∈ *set ps*. *besitz p* = *opfer-nach-besitz*}›

lemma *opfer-eindeutig-nach-besitz-auswaehlen-the-single-elem-enumall*:
‹*opfer-eindeutig-nach-besitz-auswaehlen opfer-nach-besitz besitz enum-class.enum* =
the-single-elem {*p*. *besitz p* = *opfer-nach-besitz*}›

definition *aufsummieren* :: ‹(*'person*::*enum* ⇒ *int*) ⇒ *int*› **where**
‹*aufsummieren besitz* = *sum-list* (*map besitz Enum.enum*)›

lemma ‹*aufsummieren* (*besitz* :: *person*⇒*int*) = ($\sum$ *p*←[*Alice*,*Bob*,*Carol*,*Eve*]. *besitz p*)›

lemma ‹*aufsummieren* 🌍[*Alice* := *4*, *Carol* := *8*] = *12*›
lemma ‹*aufsummieren* 🌍[*Alice* := *4*, *Carol* := *4*] = *8*›

lemma *aufsummieren-swap*:

‹*aufsummieren* (*swap p1 p2 welt*) = *aufsummieren welt*›

12 Beispiel: Zahlenwelt

Wir nehmen an, die Welt lässt sich durch eine Zahl darstellen, die den Besitz einer Person modelliert. Der Besitz ist als ganze Zahl *int* modelliert und kann auch beliebig negativ werden.

datatype *zahlenwelt* = *Zahlenwelt*
‹*person* ⇒ *int* — besitz: Besitz jeder Person.›

fun *gesamtbesitz* :: ‹*zahlenwelt* ⇒ *int*› **where**
‹*gesamtbesitz* (*Zahlenwelt besitz*) = *aufsummieren besitz*›

Beispiel:

lemma ‹*gesamtbesitz* (*Zahlenwelt* 🌍[*Alice* := *4*, *Carol* := *8*]) = *12*›
lemma ‹*gesamtbesitz* (*Zahlenwelt* 🌍[*Alice* := *4*, *Carol* := *4*]) = *8*›

Mein persönlicher Besitz:

fun *meins* :: ‹*person* ⇒ *zahlenwelt* ⇒ *int*› **where**
‹*meins p* (*Zahlenwelt besitz*) = *besitz p*›

Beispiel:

lemma ‹*meins Carol* (*Zahlenwelt* 🌍[*Alice* := *8*, *Carol* := *4*]) = *4*›

Um den `SchleierNichtwissen.thy` zu implementieren:

fun *zahlenwps* :: ‹*person* ⇒ *person* ⇒ *zahlenwelt* ⇒ *zahlenwelt*› **where**
‹*zahlenwps p1 p2* (*Zahlenwelt besitz*) = *Zahlenwelt* (*swap p1 p2 besitz*)›

Beispiel:

lemma ‹*zahlenwps Alice Carol* (*Zahlenwelt* 🌍[*Alice* := *4*, *Bob* := *6*, *Carol* := *8*])
= (*Zahlenwelt* 🌍[*Alice* := *8*, *Bob* := *6*, *Carol* := *4*])›

Alice hat Besitz, *Bob* ist reicher, *Carol* hat Schulden.

definition ‹*initialwelt* ≡ *Zahlenwelt* 🌍[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*]›

12.1 Ungültige Handlung

Sobald ich eine konkrete Person in einer Handlungsabsicht hardcode, ist diese nicht mehr wohlgeformt.

lemma ‹¬*wohlgeformte-handlungsabsicht*
zahlenwps (*Zahlenwelt* 🌍[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*])
(*Handlungsabsicht* (λ*ich w*. *if ich* = *Alice then w else Zahlenwelt* (λ-. *0*)))›

12.2 Nicht-Wohlgeformte Handlungen

```
fun stehlen :: ⟨int ⇒ person ⇒ person ⇒ zahlenwelt ⇒ zahlenwelt⟩ where
  ⟨stehlen beute opfer dieb (Zahlenwelt besitz) =
      Zahlenwelt (besitz(opfer −= beute)(dieb += beute))⟩
```

Die Handlung *stehlen* diskriminiert und ist damit nicht wohlgeformt:

```
lemma ⟨wohlgeformte-handlungsabsicht-gegenbeispiel zahlenwps
    (Zahlenwelt (λx. 0)) (Handlungsabsicht (stehlen 5 Bob))
    Alice Bob⟩
```

Wir versuchen, das Opfer nach Besitz auszuwählen, nicht nach Namen. Nach unserer Definition ist der Besitz ein Merkmal, nach dem man diskriminieren darf. Man darf nur nicht nach Eigenschaften der *person* diskriminieren, sondern nur nach Eigenschaften der *zahlenwelt*.

```
fun opfer-nach-besitz-auswaehlen :: ⟨int ⇒ ('person ⇒ int) ⇒ 'person list ⇒ 'person option⟩ where
  ⟨opfer-nach-besitz-auswaehlen - - [] = None⟩
| ⟨opfer-nach-besitz-auswaehlen b besitz (p#ps) =
    (if besitz p = b then Some p else opfer-nach-besitz-auswaehlen b besitz ps)⟩

fun stehlen2 :: ⟨int ⇒ int ⇒ person ⇒ zahlenwelt ⇒ zahlenwelt⟩ where
    ⟨stehlen2 beute opfer-nach-besitz dieb (Zahlenwelt besitz) =
      (case opfer-nach-besitz-auswaehlen opfer-nach-besitz besitz Enum.enum
         of None ⇒ (Zahlenwelt besitz)
          | Some opfer ⇒ Zahlenwelt (besitz(opfer −= beute)(dieb += beute))
      )⟩
```

Leider ist diese Funktion auch diskriminierend: Wenn es mehrere potenzielle Opfer mit dem gleichen Besitz gibt, dann bestimmt die Reihenfolge in *enum-class.enum* wer bestohlen wird. Diese Reihenfolge ist wieder eine Eigenschaft von *person* und nicht *zahlenwelt*.

```
lemma⟨handeln Alice (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := −3])
            (Handlungsabsicht (stehlen2 5 10))
= Handlung (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := − 3])
            (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := − 3])⟩
lemma⟨handeln Bob (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := −3])
            (Handlungsabsicht (stehlen2 5 10))
= Handlung (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := − 3])
          (Zahlenwelt 🌍[Alice := 5, Bob := 15, Carol := − 3])⟩
lemma ⟨wohlgeformte-handlungsabsicht-gegenbeispiel
    zahlenwps
    (Zahlenwelt 🌍[Alice := 10, Bob := 10, Carol := −3]) (Handlungsabsicht (stehlen2 5 10))
    Alice Bob⟩
```

```
fun schenken :: ⟨int ⇒ person ⇒ person ⇒ zahlenwelt ⇒ zahlenwelt⟩ where
  ⟨schenken betrag empfaenger schenker (Zahlenwelt besitz) =
      Zahlenwelt (besitz(schenker −= betrag)(empfaenger += betrag))⟩
```

Da wir ganze Zahlen verwenden und der Besitz auch beliebig negativ werden kann, ist Stehlen äquivalent dazu einen negativen Betrag zu verschenken:

lemma *stehlen-ist-schenken*: ‹*stehlen i* = *schenken* (−*i*)›

Das Modell ist nicht ganz perfekt, ... Aber passt schon um damit zu spielen.

12.3 Wohlgeformte Handlungen

Die folgende Handlung erschafft neuen Besitz aus dem Nichts:

fun *erschaffen* :: ‹*nat* ⇒ *person* ⇒ *zahlenwelt* ⇒ *zahlenwelt*› **where**
‹*erschaffen i p* (*Zahlenwelt besitz*) = *Zahlenwelt* (*besitz*(*p* += *int i*))›
lemma ‹*wohlgeformte-handlungsabsicht zahlenwps welt* (*Handlungsabsicht* (*erschaffen n*))›

Wenn wir das Opfer eindeutig auswählen, ist die Handlung wohlgeformt. Allerdings wird niemand bestohlen, wenn das Opfer nicht eindeutig ist.

fun *stehlen4* :: ‹*int* ⇒ *int* ⇒ *person* ⇒ *zahlenwelt* ⇒ *zahlenwelt*› **where**
‹*stehlen4 beute opfer-nach-besitz dieb* (*Zahlenwelt besitz*) =
(*case opfer-eindeutig-nach-besitz-auswaehlen opfer-nach-besitz besitz Enum.enum*
of None ⇒ (*Zahlenwelt besitz*)
| *Some opfer* ⇒ *Zahlenwelt* (*besitz*(*opfer* −= *beute*)(*dieb* += *beute*))
)›

Reset versetzt die Welt wieder in den Ausgangszustand. Eine sehr destruktive Handlung.

fun *reset* :: ‹*person* ⇒ *zahlenwelt* ⇒ *zahlenwelt*› **where**
‹*reset ich* (*Zahlenwelt besitz*) = *Zahlenwelt* (λ -. *0*)›

Der *reset* ist im moralischen Sinne vermutlich keine gute Handlung, dennoch ist es eine wohlgeformte Handlung, welche wir betrachten können:

lemma ‹*wohlgeformte-handlungsabsicht zahlenwps welt* (*Handlungsabsicht reset*)›

fun *alles-kaputt-machen* :: ‹*person* ⇒ *zahlenwelt* ⇒ *zahlenwelt*› **where**
‹*alles-kaputt-machen ich* (*Zahlenwelt besitz*) = *Zahlenwelt* (λ -. *Min* (*besitz* ‘ *UNIV*) − *1*)›

lemma ‹*alles-kaputt-machen Alice* (*Zahlenwelt* 🌏[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*])
= (*Zahlenwelt* 🌏[*Alice* := −*4*, *Bob* := −*4*, *Carol* := −*4*, *Eve* := −*4*])›

Die Beispielhandlungsabsichten, die wir betrachten wollen.

definition *handlungsabsichten* ≡ [
Handlungsabsicht (*erschaffen 5*),
Handlungsabsicht (*stehlen4 5 10*),
Handlungsabsicht reset,
Handlungsabsicht alles-kaputt-machen
]

lemma ‹*ha* ∈ *set handlungsabsichten* ⟹ *wohlgeformte-handlungsabsicht zahlenwps welt ha*›

12.4 Maxime für individuellen Fortschritt

Wir definieren eine Maxime die besagt, dass sich der Besitz einer Person nicht verringern darf:

fun *individueller-fortschritt* :: ‹*person* ⇒ *zahlenwelt handlung* ⇒ *bool*› **where**
‹*individueller-fortschritt p* (*Handlung vor nach*) ⟷ (*meins p vor*) ≤ (*meins p nach*)›

definition *maxime-zahlenfortschritt* :: ‹(*person*, *zahlenwelt*) *maxime*› **where**
‹*maxime-zahlenfortschritt* ≡ *Maxime* (λ*ich*. *individueller-fortschritt ich*)›

reset erfüllt das nicht, aber das normale *stehlen*.

lemma *ha* ∈ {
Handlungsabsicht (*erschaffen 5*),
Handlungsabsicht (*stehlen 5 Bob*),
Handlungsabsicht (*stehlen4 5 10*),
Handlungsabsicht alles-kaputt-machen
} ⟹ *maxime-und-handlungsabsicht-generalisieren zahlenwps welt maxime-zahlenfortschritt ha p*

Gilt nicht:

lemma
‹*maxime-und-handlungsabsicht-generalisieren zahlenwps welt*
maxime-zahlenfortschritt (*Handlungsabsicht* (*reset*)) *p*›

Die *maxime-zahlenfortschritt* erfüllt **nicht** den *kategorischer-imperativ* da *Alice* nach der Maxime z.B. *Bob* bestehlen dürfte.

lemma ‹*kategorischer-imperativ-gegenbeispiel*
zahlenwps initialwelt maxime-zahlenfortschritt
(*Handlungsabsicht* (*stehlen4 1 10*))
Alice
Bob
Alice›

12.4.1 Einzellbeispiele

In jeder Welt ist die *Handlungsabsicht* (*erschaffen n*) *moralisch*:

lemma ‹*moralisch welt maxime-zahlenfortschritt* (*Handlungsabsicht* (*erschaffen n*))›

In kein Welt ist Stehlen *moralisch*:

lemma ‹¬ *moralisch welt maxime-zahlenfortschritt* (*Handlungsabsicht* (*stehlen 5 Bob*))›

In unserer *initialwelt* in der *Bob* als Opfer anhand seines Besitzes als Opfer eines Diebstahls ausgewählt würde, ist stehlen dennoch nicht *moralisch*, obwohl die Handlungsabsicht wohlgeformt ist:

lemma ‹¬ *moralisch initialwelt maxime-zahlenfortschritt* (*Handlungsabsicht* (*stehlen4 5 10*))›

Da Schenken und Stehlen in dieser Welt equivalent ist, ist Schenken auch unmoralisch:

lemma ‹¬ *moralisch welt maxime-zahlenfortschritt* (*Handlungsabsicht* (*schenken 5 Bob*))›

TODO: erklaeren

lemma ‹*erzeuge-beispiel*
zahlenwps initialwelt
handlungsabsichten
(*Maxime individueller-fortschritt*) =
Some
⦇*bsp-welt* = *Zahlenwelt* 🌏[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*],
bsp-erfuellte-maxime = *None*,
bsp-erlaubte-handlungen = [*Handlungsabsicht* (*erschaffen 5*)],
bsp-verbotene-handlungen = [*Handlungsabsicht* (*stehlen4 5 10*), *Handlungsabsicht reset*, *Handlungsabsicht alles-kaputt-machen*]⦈›

12.5 Maxime für allgemeinen Fortschritt

Allerdings können wir die Maxime generalisieren, indem wir *individueller-fortschritt* für jeden fordern. Effektiv wird dabei das *ich* ignoriert.

definition *maxime-altruistischer-fortschritt* :: ‹(*person*, *zahlenwelt*) *maxime*› **where**
‹*maxime-altruistischer-fortschritt* ≡ *Maxime* (λ*ich h*. ∀ *pX*. *individueller-fortschritt pX h*)›

Folgendes Beispiel zeigt, dass die *maxime-altruistischer-fortschritt* den kategorischen Imperativ (für diese *initialwelt* und *handlungsabsichten*) erfüllt; zu sehen an dem *Some* Term im *bsp-erfuellte-maxime*. Die Handlungsabsichten werden eingeordnet wie erwartet: *erschaffen* ist gut, *stehlen4*, *reset*, *alles-kaputt-machen* ist schlecht.

lemma ‹*erzeuge-beispiel*
zahlenwps initialwelt
handlungsabsichten
maxime-altruistischer-fortschritt =
Some
⦇*bsp-welt* = *Zahlenwelt* 🌏[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*],
bsp-erfuellte-maxime = *Some maxime-altruistischer-fortschritt*,
bsp-erlaubte-handlungen = [*Handlungsabsicht* (*erschaffen 5*)],
bsp-verbotene-handlungen = [*Handlungsabsicht* (*stehlen4 5 10*), *Handlungsabsicht reset*, *Handlungsabsicht alles-kaputt-machen*]⦈›

Das ist ein sehr schönes Beispiel.

Die Aussage, dass die *maxime-altruistischer-fortschritt* den kategorischen Imperativ für bestimmte Handlungsabsichten und Welten erfüllt generalisiert noch weiter. Für alle Welten und alle wohlgeformten Handlungsabsichten welche mit der Maxime generalisieren erfüllt die Maxime den kategorischen Imperativ.

theorem ‹
  ∀p. maxime-und-handlungsabsicht-generalisieren zahlenwps welt maxime-altruistischer-fortschritt ha p ⟹
  wohlgeformte-handlungsabsicht zahlenwps welt ha ⟹
  kategorischer-imperativ-auf ha welt maxime-altruistischer-fortschritt›

Allgemein scheint dies eine sehr gute Maxime zu sein (für dieses sehr beschränkte Weltenmodell).

12.6 Maxime für strikten individuellen Fortschritt

In der Maxime *individueller-fortschritt* hatten wir *meins p vor* $\leq$ *meins p nach*. Was wenn wir nun echten Fortschritt fordern: *meins p vor* $<$ *meins p nach*.

fun *individueller-strikter-fortschritt* :: ‹*person ⇒ zahlenwelt handlung ⇒ bool*› **where**
  ‹*individueller-strikter-fortschritt p (Handlung vor nach) ⟷ (meins p vor) < (meins p nach)*›

TODO: erklaeren. Erfuellt nicht kategorischen imperativ und alles ist verboten

lemma ‹*erzeuge-beispiel*
  zahlenwps initialwelt
  handlungsabsichten
  (Maxime individueller-strikter-fortschritt) =
Some
  ⦇bsp-welt = Zahlenwelt 🌐[Alice := 5, Bob := 10, Carol := −3],
   bsp-erfuellte-maxime = None,
   bsp-erlaubte-handlungen = [],
   bsp-verbotene-handlungen = handlungsabsichten⦈›

In keiner Welt ist die Handlung *erschaffen* nun *moralisch*:

lemma ‹*¬ moralisch welt*
    (Maxime (λich. individueller-strikter-fortschritt ich)) (Handlungsabsicht (erschaffen 5))›

Der Grund ist, dass der Rest der Bevölkerung keine *strikte* Erhöhung des eigenen Wohlstands erlebt. Effektiv führt diese Maxime zu einem Gesetz, welches es einem Individuum nicht erlaubt mehr Besitz zu erschaffen, obwohl niemand dadurch einen Nachteil hat. Diese Maxime kann meiner Meinung nach nicht gewollt sein.

Beispielsweise ist *Bob* das Opfer wenn *Alice* sich 5 Wohlstand erschafft, aber *Bob*'s Wohlstand sich nicht erhöht:

lemma‹*VerletzteMaxime (Opfer Bob) (Taeter Alice)*
    (Handlung [(Alice, 5), (Bob, 10), (Carol, −3)] [(Alice, 10), (Bob, 10), (Carol, −3)])
    ∈ debug-maxime show-zahlenwelt initialwelt
     (Maxime (λich. individueller-strikter-fortschritt ich)) (Handlungsabsicht (erschaffen 5)) ›

12.7 Maxime für globales striktes Optimum

Wir bauen nun eine Maxime, die das Individuum vernachlässigt und nur nach dem globalen Optimum strebt:

```
fun globaler-strikter-fortschritt :: ‹zahlenwelt handlung ⇒ bool› where
  ‹globaler-strikter-fortschritt (Handlung vor nach) ⟷ (gesamtbesitz vor) < (gesamtbesitz nach)›
```

Die Maxime ignoriert das *ich* komplett.

Nun ist es *Alice* wieder erlaubt, Wohlstand für sich selbst zu erzeugen, da sich dadurch auch der Gesamtwohlstand erhöht:

```
lemma ‹moralisch initialwelt
        (Maxime (λich. globaler-strikter-fortschritt)) (Handlungsabsicht (erschaffen 5))›
```

Allerdings ist auch diese Maxime auch sehr grausam, da sie Untätigkeit verbietet:

```
lemma ‹¬moralisch initialwelt
        (Maxime (λich. globaler-strikter-fortschritt)) (Handlungsabsicht (erschaffen 0))›
```

Unsere initiale einfache *maxime-zahlenfortschritt* würde Untätigkeit hier erlauben:

```
lemma ‹moralisch initialwelt
        maxime-zahlenfortschritt (Handlungsabsicht (erschaffen 0))›
```

TODO: erklaeren.

```
lemma ‹erzeuge-beispiel
  zahlenwps initialwelt
  handlungsabsichten
  (Maxime (λich. globaler-strikter-fortschritt)) =
Some
  (|bsp-welt = Zahlenwelt 🌏[Alice := 5, Bob := 10, Carol := −3],
   bsp-erfuellte-maxime = Some (Maxime (λich. globaler-strikter-fortschritt)),
   bsp-erlaubte-handlungen = [Handlungsabsicht (erschaffen 5)],
   bsp-verbotene-handlungen = [
    Handlungsabsicht (stehlen4 5 10),
    Handlungsabsicht reset,
    Handlungsabsicht alles-kaputt-machen]|)›
```

12.8 Maxime für globales Optimum

Wir können die Maxime für globalen Fortschritt etwas lockern:

```
fun globaler-fortschritt :: ‹zahlenwelt handlung ⇒ bool› where
 ‹globaler-fortschritt (Handlung vor nach) ⟷ (gesamtbesitz vor) ≤ (gesamtbesitz nach)›
```

Untätigkeit ist nun auch hier erlaubt:

```
lemma ‹moralisch initialwelt
```

(*Maxime* (*λich. globaler-fortschritt*)) (*Handlungsabsicht* (*erschaffen 0*))›
theorem
‹∀ *p. maxime-und-handlungsabsicht-generalisieren zahlenwps welt*
(*Maxime* (*λich. globaler-fortschritt*)) *ha p* ⟹
wohlgeformte-handlungsabsicht zahlenwps welt ha ⟹
kategorischer-imperativ-auf ha welt (*Maxime* (*λich::person. globaler-fortschritt*))›

Allerdings ist auch Stehlen erlaubt, da global gesehen, kein Besitz vernichtet wird:

lemma ‹*moralisch initialwelt*
(*Maxime* (*λich. globaler-fortschritt*)) (*Handlungsabsicht* (*stehlen 5 Bob*))›
lemma ‹*moralisch initialwelt*
(*Maxime* (*λich. globaler-fortschritt*)) (*Handlungsabsicht* (*stehlen4 5 10*))›

TODO: erklaeren.

lemma ‹*erzeuge-beispiel*
zahlenwps initialwelt
handlungsabsichten
(*Maxime* (*λich. globaler-fortschritt*)) =
Some
⦇*bsp-welt* = *Zahlenwelt* 🌏[*Alice* := *5*, *Bob* := *10*, *Carol* := −*3*],
bsp-erfuellte-maxime = *Some* (*Maxime* (*λich. globaler-fortschritt*)),
bsp-erlaubte-handlungen = [*Handlungsabsicht* (*erschaffen 5*), *Handlungsabsicht* (*stehlen4 5 10*)],
bsp-verbotene-handlungen = [
Handlungsabsicht reset,
Handlungsabsicht alles-kaputt-machen]⦈›

12.9 Ungültige Maxime

Es ist verboten, in einer Maxime eine spezielle Person hardzucoden. Da dies gegen die Gleichbehandlung aller Menschen verstoßen würde.

Beispielsweise könnten wir *individueller-fortschritt* nicht mehr parametrisiert verwenden, sondern einfach *Alice* reinschreiben:

lemma ‹*individueller-fortschritt Alice*
= (*λh. case h of Handlung vor nach* ⇒ (*meins Alice vor*) ≤ (*meins Alice nach*))›
lemma ¬*wohlgeformte-maxime-auf*
(*handeln Alice initialwelt* (*Handlungsabsicht* (*stehlen4 5 10*))) *zahlenwps*
(*Maxime* (*λich. individueller-fortschritt Alice*))
lemma *wohlgeformte-maxime-auf*
(*handeln Alice initialwelt* (*Handlungsabsicht* (*stehlen4 5 10*))) *zahlenwps*
(*Maxime* (*λich. individueller-fortschritt ich*))

13 Gesetz

Definiert einen Datentyp um Gesetzestext zu modellieren.

```
datatype 'a tatbestand = Tatbestand ‹'a›

datatype 'a rechtsfolge = Rechtsfolge ‹'a›

datatype ('a, 'b) rechtsnorm = Rechtsnorm ‹'a tatbestand› ‹'b rechtsfolge›

datatype 'p prg = Paragraph ‹'p› (§)

datatype ('p, 'a, 'b) gesetz = Gesetz ‹('p prg × ('a, 'b) rechtsnorm) set›
```

Beispiel, von https://de.wikipedia.org/wiki/Rechtsfolge:

```
value ‹Gesetz {
  (§ ''823 BGB'',
   Rechtsnorm
     (Tatbestand ''Wer vorsaetzlich oder fahrlaessig das Leben, den Koerper, die Gesundheit, (...),
                 das Eigentum oder (...) eines anderen widerrechtlich verletzt,'')
     (Rechtsfolge ''ist dem anderen zum Ersatz des daraus entstehenden Schadens verpflichtet.'')
  ),
  (§ ''985 BGB'',
   Rechtsnorm
     (Tatbestand ''Der Eigentuemer einer Sache kann von dem Besitzer'')
     (Rechtsfolge ''die Herausgabe der Sache verlangen'')
  ),
  (§ ''303 StGB'',
   Rechtsnorm
     (Tatbestand ''Wer rechtswidrig eine fremde Sache beschaedigt oder zerstoert,'')
     (Rechtsfolge ''wird mit Freiheitsstrafe bis zu zwei Jahren oder mit Geldstrafe bestraft.'')
  )
  }›


fun neuer-paragraph :: ‹(nat, 'a, 'b) gesetz ⇒ nat prg› where
 ‹neuer-paragraph (Gesetz G) = § ((max-paragraph (fst ` G)) + 1)›
```

Fügt eine Rechtsnorm als neuen Paragraphen hinzu:

```
fun hinzufuegen :: ‹('a,'b) rechtsnorm ⇒ (nat,'a,'b) gesetz ⇒ (nat,'a,'b) gesetz› where
  ‹hinzufuegen rn (Gesetz G) =
    (if rn ∈ (snd ` G) then Gesetz G else Gesetz (insert (neuer-paragraph (Gesetz G), rn) G))›
```

Modelliert ob eine Handlung ausgeführt werden muss, darf, kann, nicht muss:

```
datatype sollensanordnung = Gebot | Verbot | Erlaubnis | Freistellung
```

Beispiel:

```
lemma ‹hinzufuegen
        (Rechtsnorm (Tatbestand ''tb2'') (Rechtsfolge Verbot))
```

(*Gesetz* {(§ *1*, (*Rechtsnorm* (*Tatbestand* ''*tb1*'') (*Rechtsfolge Erlaubnis*)))}) =
Gesetz
{(§ *2*, *Rechtsnorm* (*Tatbestand* ''*tb2*'') (*Rechtsfolge Verbot*)),
(§ *1*, *Rechtsnorm* (*Tatbestand* ''*tb1*'') (*Rechtsfolge Erlaubnis*))}›

14 Experimental: Moralisch Gesetzs Ableiten

14.1 Allgemeines Gesetz Ableiten

Wir wollen implementieren:
„Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein ***allgemeines Gesetz*** *werde.“*
Für eine gebene Welt haben wir schon eine Handlung nach einer Maxime untersucht: *moralisch*
Das Ergebnis sagt uns ob diese Handlung gut oder schlecht ist. Basierend darauf müssen wir nun ein allgemeines Gesetz ableiten.
Ich habe keine Ahnung wie das genau funktionieren soll, deswegen schreibe ich einfach nur in einer Typsignatur auf, was zu tun ist:
Gegeben:

- *'world handlung*: Die Handlung
- *sollensanordnung*: Das Ergebnis der moralischen Bewertung, ob die Handlung gut/schlecht.

Gesucht:

- (*'a*, *'b*) *rechtsnorm*: ein allgemeines Gesetz

type-synonym (*'world*, *'a*, *'b*) *allgemeines-gesetz-ableiten* =
‹*'world handlung* ⇒ *sollensanordnung* ⇒ (*'a*, *'b*) *rechtsnorm*›

Soviel vorweg: Nur aus einer von außen betrachteten Handlung und einer Entscheidung ob diese Handlung ausgeführt werden soll wird es schwer ein allgemeines Gesetz abzuleiten.

14.2 Implementierung Moralisch ein Allgemeines Gesetz Ableiten

Und nun werfen wir alles zusammen:
„Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.“
Eingabe:

- *'person*: handelnde Person
- *'world*: Die Welt in ihrem aktuellen Zustand

- ('*person*, '*world*) *handlungsabsicht*: Eine mögliche Handlung, über die wir entscheiden wollen ob wir sie ausführen sollten.
- ('*person*, '*world*) *maxime*: Persönliche Ethik.
- ('*world*, '*a*, '*b*) *allgemeines-gesetz-ableiten*: wenn man keinen Plan hat wie man sowas implementiert, einfach als Eingabe annehmen.
- (*nat*, '*a*, '*b*) *gesetz*: Initiales allgemeines Gesetz (normalerweise am Anfang leer).

Ausgabe: *sollensanordnung*: Sollen wir die Handlung ausführen? (*nat*, '*a*, '*b*) *gesetz*: Soll das allgemeine Gesetz entsprechend angepasst werden?

```
definition moarlisch-gesetz-ableiten ::
  ‹'person ⇒
   'world ⇒
   ('person, 'world) maxime ⇒
   ('person, 'world) handlungsabsicht ⇒
   ('world, 'a, 'b) allgemeines-gesetz-ableiten ⇒
   (nat, 'a, 'b) gesetz
  ⇒ (sollensanordnung × (nat, 'a, 'b) gesetz)›
where
  ‹moarlisch-gesetz-ableiten ich welt maxime handlungsabsicht gesetz-ableiten gesetz ≡
    let soll-handeln = if moralisch welt maxime handlungsabsicht
                       then
                         Erlaubnis
                       else
                         Verbot in
     (
       soll-handeln,
       hinzufuegen (gesetz-ableiten (handeln ich welt handlungsabsicht) soll-handeln) gesetz
     )›
```

Das ganze *moarlisch-gesetz-ableiten* dient mehr dem Debugging, ...

15 Gesetze

Wir implementieren Strategien um ('*world*, '*a*, '*b*) *allgemeines-gesetz-ableiten* zu implementieren.

15.1 Case Law Absolut

Gesetz beschreibt: wenn (vorher, nachher) dann Erlaubt/Verboten, wobei vorher/nachher die Welt beschreiben. Paragraphen sind einfache natürliche Zahlen.

```
type-synonym 'world case-law = ‹(nat, ('world × 'world), sollensanordnung) gesetz›
```

Überträgt einen Tatbestand wörtlich ins Gesetz. Nicht sehr allgemein.

```
definition case-law-ableiten-absolut
```

```
  :: ‹('world, ('world × 'world), sollensanordnung) allgemeines-gesetz-ableiten›
  where
    ‹case-law-ableiten-absolut handlung sollensanordnung =
        Rechtsnorm
          (Tatbestand (vorher handlung, nachher handlung))
          (Rechtsfolge sollensanordnung)›
```

```
definition printable-case-law-ableiten-absolut
  :: ‹('world ⇒'printable-world) ⇒
    ('world, ('printable-world × 'printable-world), sollensanordnung) allgemeines-gesetz-ableiten›
  where
  ‹printable-case-law-ableiten-absolut print-world h ≡
      case-law-ableiten-absolut (map-handlung print-world h)›
```

15.2 Case Law Relativ

Case Law etwas besser, wir zeigen nur die Änderungen der Welt.

```
fun case-law-ableiten-relativ
    :: ‹('world handlung ⇒ (('person, 'etwas) aenderung) list)
        ⇒ ('world, (('person, 'etwas) aenderung) list, sollensanordnung)
            allgemeines-gesetz-ableiten›
  where
    ‹case-law-ableiten-relativ delta handlung erlaubt =
      Rechtsnorm (Tatbestand (delta handlung)) (Rechtsfolge erlaubt)›
```

16 Simulation

Gegeben eine handelnde Person und eine Maxime, wir wollen simulieren was für ein allgemeines Gesetz abgeleitet werden könnte.

```
datatype ('person, 'world, 'a, 'b) simulation-constants = SimConsts
    ‹'person› — handelnde Person
    ‹('person, 'world) maxime›
    ‹('world, 'a, 'b) allgemeines-gesetz-ableiten›
```

...

... Die Funktion *simulateOne* nimmt eine Konfiguration (*'person*, *'world*, *'a*, *'b*) *simulation-constants*, eine Anzahl an Iterationen die durchgeführt werden sollen, eine Handlung, eine Initialwelt, ein Initialgesetz, und gibt das daraus resultierende Gesetz nach so vielen Iterationen zurück.

Beispiel: Wir nehmen die mir-ist-alles-egal Maxime. Wir leiten ein allgemeines Gesetz ab indem wir einfach nur die Handlung wörtlich ins Gesetz übernehmen. Wir machen *10::'a* Iterationen. Die Welt ist nur eine Zahl und die initiale Welt sei *32::'a*. Die Handlung ist es diese Zahl um Eins zu erhöhen, Das Ergebnis der Simulation ist dann, dass wir einfach von *32::'a* bis *42::'a* zählen.

```
lemma ‹simulateOne
      (SimConsts () (Maxime (λ- -. True)) (λh s. Rechtsnorm (Tatbestand h) (Rechtsfolge ''count'')))
```

10 (*Handlungsabsicht* (λp *n*. *Suc n*))
32
(*Gesetz* {}) =
Gesetz
{(§ *10*, *Rechtsnorm* (*Tatbestand* (*Handlung 41 42*)) (*Rechtsfolge ''count''*)),
(§ *9*, *Rechtsnorm* (*Tatbestand* (*Handlung 40 41*)) (*Rechtsfolge ''count''*)),
(§ *8*, *Rechtsnorm* (*Tatbestand* (*Handlung 39 40*)) (*Rechtsfolge ''count''*)),
(§ *7*, *Rechtsnorm* (*Tatbestand* (*Handlung 38 39*)) (*Rechtsfolge ''count''*)),
(§ *6*, *Rechtsnorm* (*Tatbestand* (*Handlung 37 38*)) (*Rechtsfolge ''count''*)),
(§ *5*, *Rechtsnorm* (*Tatbestand* (*Handlung 36 37*)) (*Rechtsfolge ''count''*)),
(§ *4*, *Rechtsnorm* (*Tatbestand* (*Handlung 35 36*)) (*Rechtsfolge ''count''*)),
(§ *3*, *Rechtsnorm* (*Tatbestand* (*Handlung 34 35*)) (*Rechtsfolge ''count''*)),
(§ *2*, *Rechtsnorm* (*Tatbestand* (*Handlung 33 34*)) (*Rechtsfolge ''count''*)),
(§ *1*, *Rechtsnorm* (*Tatbestand* (*Handlung 32 33*)) (*Rechtsfolge ''count''*))}›

Eine Iteration der Simulation liefert genau einen Paragraphen im Gesetz:

lemma ‹∃ *tb rf*.
simulateOne
(*SimConsts person maxime gesetz-ableiten*)
1 handlungsabsicht
initialwelt
(*Gesetz* {})
= *Gesetz* {(§ *1*, *Rechtsnorm* (*Tatbestand tb*) (*Rechtsfolge rf*))}›

17 Beispiel: BeispielZahlenwelt aber mit Gesetz (Experimental)

17.1 Setup

Wir nehmen an unsere handelnde Person ist *Alice*.

definition ‹*beispiel-case-law-absolut maxime handlungsabsicht* ≡
simulateOne
(*SimConsts*
Alice
maxime
(*printable-case-law-ableiten-absolut show-zahlenwelt*))
5 handlungsabsicht initialwelt (*Gesetz* {})›
definition ‹*beispiel-case-law-relativ maxime handlungsabsicht* ≡
simulateOne
(*SimConsts*
Alice
maxime
(*case-law-ableiten-relativ delta-zahlenwelt*))
10 handlungsabsicht initialwelt (*Gesetz* {})›

17.2 Beispiele

Alice kann beliebig oft 5 Wohlstand für sich selbst erschaffen. Das entstehende Gesetz ist nicht sehr gut, da es einfach jedes Mal einen Snapshot der Welt aufschreibt und nicht sehr generisch ist.

lemma ‹*beispiel-case-law-absolut maxime-zahlenfortschritt* (*Handlungsabsicht* (*erschaffen 5*))
=
Gesetz
{(§ *5*,
Rechtsnorm
(*Tatbestand* ([(*Alice*, *25*), (*Bob*, *10*), (*Carol*, − *3*)], [(*Alice*, *30*), (*Bob*, *10*), (*Carol*, − *3*)]))
(*Rechtsfolge Erlaubnis*)),
(§ *4*,
Rechtsnorm
(*Tatbestand* ([(*Alice*, *20*), (*Bob*, *10*), (*Carol*, − *3*)], [(*Alice*, *25*), (*Bob*, *10*), (*Carol*, − *3*)]))
(*Rechtsfolge Erlaubnis*)),
(§ *3*,
Rechtsnorm
(*Tatbestand* ([(*Alice*, *15*), (*Bob*, *10*), (*Carol*, − *3*)], [(*Alice*, *20*), (*Bob*, *10*), (*Carol*, − *3*)]))
(*Rechtsfolge Erlaubnis*)),
(§ *2*,
Rechtsnorm
(*Tatbestand* ([(*Alice*, *10*), (*Bob*, *10*), (*Carol*, − *3*)], [(*Alice*, *15*), (*Bob*, *10*), (*Carol*, − *3*)]))
(*Rechtsfolge Erlaubnis*)),
(§ *1*,
Rechtsnorm
(*Tatbestand* ([(*Alice*, *5*), (*Bob*, *10*), (*Carol*, − *3*)], [(*Alice*, *10*), (*Bob*, *10*), (*Carol*, − *3*)]))
(*Rechtsfolge Erlaubnis*))}
›

Die gleiche Handlung, wir schreiben aber nur die Änderung der Welt ins Gesetz:

lemma ‹*beispiel-case-law-relativ maxime-zahlenfortschritt* (*Handlungsabsicht* (*erschaffen 5*)) =
Gesetz
{(§ *1*, *Rechtsnorm* (*Tatbestand* [*Gewinnt Alice 5*]) (*Rechtsfolge Erlaubnis*))}›

lemma ‹*beispiel-case-law-relativ*
(*Maxime* (λ(*ich*::*person*) *h*. (∀ *pX*. *individueller-fortschritt pX h*))) (*Handlungsabsicht* (*erschaffen 5*)) =
Gesetz {(§ *1*, *Rechtsnorm* (*Tatbestand* [*Gewinnt Alice 5*]) (*Rechtsfolge Erlaubnis*))}›

Nun ist es *Alice* verboten Wohlstand für sich selbst zu erzeugen.

lemma ‹*beispiel-case-law-relativ*
(*Maxime* (λ*ich*. *individueller-strikter-fortschritt ich*))
(*Handlungsabsicht* (*erschaffen 5*)) =
Gesetz {(§ *1*, *Rechtsnorm* (*Tatbestand* [*Gewinnt Alice 5*]) (*Rechtsfolge Verbot*))}›

lemma ‹*beispiel-case-law-relativ*
(Maxime (λich. globaler-strikter-fortschritt))
(Handlungsabsicht (erschaffen 5)) =
Gesetz {(§ 1, Rechtsnorm (Tatbestand [Gewinnt Alice 5]) (Rechtsfolge Erlaubnis))}›

lemma ‹*beispiel-case-law-relativ*
(Maxime (λich. globaler-strikter-fortschritt))
(Handlungsabsicht (erschaffen 0)) =
Gesetz {(§ 1, Rechtsnorm (Tatbestand []) (Rechtsfolge Verbot))}›

lemma ‹*beispiel-case-law-relativ*
maxime-zahlenfortschritt
(Handlungsabsicht (erschaffen 0)) =
Gesetz {(§ 1, Rechtsnorm (Tatbestand []) (Rechtsfolge Erlaubnis))}›

lemma‹*beispiel-case-law-relativ*
(Maxime (λich. globaler-fortschritt))
(Handlungsabsicht (erschaffen 0))
=
Gesetz {(§ 1, Rechtsnorm (Tatbestand []) (Rechtsfolge Erlaubnis))}›

lemma‹*beispiel-case-law-relativ*
(Maxime (λich. globaler-fortschritt))
(Handlungsabsicht (stehlen 5 Bob))
=
Gesetz
{(§ 1, Rechtsnorm (Tatbestand [Gewinnt Alice 5, Verliert Bob 5]) (Rechtsfolge Erlaubnis))}›

Stehlen ist verboten:

lemma ‹*beispiel-case-law-relativ maxime-zahlenfortschritt (Handlungsabsicht (stehlen 5 Bob)) =*
Gesetz
{(§ 1, Rechtsnorm (Tatbestand [Gewinnt Alice 5, Verliert Bob 5]) (Rechtsfolge Verbot))}›

Auch wenn *Alice* von sich selbst stehlen möchte ist dies verboten, obwohl hier keiner etwas verliert:

lemma ‹*beispiel-case-law-relativ maxime-zahlenfortschritt (Handlungsabsicht (stehlen 5 Alice)) =*
Gesetz {(§ 1, Rechtsnorm (Tatbestand []) (Rechtsfolge Verbot))}›

Der Grund ist, dass *Alice* die abstrakte Handlung "Alice wird bestohlen" gar nicht gut fände, wenn sie jemand anderes ausführt:

lemma ‹*debug-maxime show-zahlenwelt initialwelt*
maxime-zahlenfortschritt (*Handlungsabsicht* (*stehlen 5 Alice*)) =
{*VerletzteMaxime* (*Opfer Alice*) (*Taeter Bob*)
(*Handlung* [(*Alice*, *5*), (*Bob*, *10*), (*Carol*, − *3*)] [(*Bob*, *15*), (*Carol*, − *3*)]),
VerletzteMaxime (*Opfer Alice*) (*Taeter Carol*)
(*Handlung* [(*Alice*, *5*), (*Bob*, *10*), (*Carol*, − *3*)] [(*Bob*, *10*), (*Carol*, *2*)]),
VerletzteMaxime (*Opfer Alice*) (*Taeter Eve*)
(*Handlung* [(*Alice*, *5*), (*Bob*, *10*), (*Carol*, − *3*)] [(*Bob*, *10*), (*Carol*, − *3*), (*Eve*, *5*)])
}›

Leider ist das hier abgeleitete Gesetz sehr fragwürdig: *Rechtsnorm* (*Tatbestand* []) (*Rechtsfolge Verbot*) Es besagt, dass Nichtstun verboten ist.

Indem wir die beiden Handlungen Nichtstun und Selbstbestehlen betrachten, können wir sogar ein widersprüchliches Gesetz ableiten:

lemma ‹*simulateOne*
(*SimConsts*
Alice
maxime-zahlenfortschritt
(*case-law-ableiten-relativ delta-zahlenwelt*))
20 (*Handlungsabsicht* (*stehlen 5 Alice*)) *initialwelt*
(*beispiel-case-law-relativ maxime-zahlenfortschritt* (*Handlungsabsicht* (*erschaffen 0*)))
=
Gesetz
{(§ *2*, *Rechtsnorm* (*Tatbestand* []) (*Rechtsfolge Verbot*)),
(§ *1*, *Rechtsnorm* (*Tatbestand* []) (*Rechtsfolge Erlaubnis*))}›

Meine persönliche Conclusion: Wir müssen irgendwie die Absicht mit ins Gesetz schreiben.

Es ist *Alice* verboten, etwas zu verschenken:

lemma‹*beispiel-case-law-relativ maxime-zahlenfortschritt* (*Handlungsabsicht* (*schenken 5 Bob*))
=
Gesetz
{(§ *1*,
Rechtsnorm (*Tatbestand* [*Verliert Alice 5*, *Gewinnt Bob 5*]) (*Rechtsfolge Verbot*))}›

Der Grund ist, dass *Alice* dabei etwas verliert und die *maxime-zahlenfortschritt* dies nicht Erlaubt. Es fehlt eine Möglichkeit zu modellieren, dass *Alice* damit einverstanden ist, etwas abzugeben. Doch wir haben bereits in *stehlen i* = *schenken* (− *i*) gesehen, dass *stehlen* und *schenken* nicht unterscheidbar sind.

Folgende ungültige Maxime würde es erlauben, dass *Alice* Leute bestehlen darf:

lemma‹*beispiel-case-law-relativ*
(*Maxime* (λ*ich*. *individueller-fortschritt Alice*))

(*Handlungsabsicht* (*stehlen 5 Bob*))
=
Gesetz
{(§ *1*, *Rechtsnorm* (*Tatbestand* [*Gewinnt Alice 5*, *Verliert Bob 5*]) (*Rechtsfolge Erlaubnis*))}›

18 Einkommensteuergesetzgebung

Basierend auf einer stark vereinfachten Version des deutschen Steuerrechts. Wenn ich Wikipedia richtig verstanden habe, habe ich sogar aus Versehen einen Teil des österreichischen Steuersystem gebaut mit deutschen Konstanten.

Folgende **locale** nimmt an, dass wir eine Funktion *steuer*::*nat* ⇒ *nat* haben, welche basierend auf dem Einkommen die zu zahlende Steuer berechnet.

Die *steuer* Funktion arbeitet auf natürlichen Zahlen. Wir nehmen an, dass einfach immer auf ganze Geldbeträge gerundet wird. Wie im deutschen System.

Die **locale** einhält einige Definition, gegeben die *steuer* Funktion.

Eine konkrete *steuer* Funktion wird noch nicht gegeben.

```
locale steuer-defs =
  fixes steuer :: ‹nat ⇒ nat› — Einkommen -> Steuer
begin
  definition brutto :: ‹nat ⇒ nat› where
    ‹brutto einkommen ≡ einkommen›
  definition netto :: ‹nat ⇒ nat› where
    ‹netto einkommen ≡ einkommen − (steuer einkommen)›
  definition steuersatz :: ‹nat ⇒ percentage› where
    ‹steuersatz einkommen ≡ percentage ((steuer einkommen) / einkommen)›
end
```

Beispiel. Die *steuer* Funktion sagt, man muss 25 Prozent Steuern zahlen:

```
definition beispiel-25prozent-steuer :: ‹nat ⇒ nat› where
  ‹beispiel-25prozent-steuer e ≡ nat ⌊real e ∗ (percentage 0.25)⌋›

lemma
  ‹beispiel-25prozent-steuer 100 = 25›
  ‹steuer-defs.brutto 100 = 100›
  ‹steuer-defs.netto beispiel-25prozent-steuer 100 = 75›
  ‹steuer-defs.steuersatz beispiel-25prozent-steuer 100 = percentage 0.25›
```

Folgende **locale** erweitert die *steuer-defs* **locale** und stellt einige Anforderungen die eine gültige *steuer* Funktion erfüllen muss.

- Wer mehr Einkommen hat, muss auch mehr Steuern zahlen.
- Leistung muss sich lohnen: Wer mehr Einkommen hat muss auch nach Abzug der Steuer mehr übrig haben.

- Existenzminimum: Es gibt ein Existenzminimum, welches nicht besteuert werden darf.

```
locale steuersystem = steuer-defs +
  assumes wer-hat-der-gibt:
    ‹einkommen-a ≥ einkommen-b ⟹ steuer einkommen-a ≥ steuer einkommen-b›

  and leistung-lohnt-sich:
    ‹einkommen-a ≥ einkommen-b ⟹ netto einkommen-a ≥ netto einkommen-b›

  — Ein Existenzminimum wird nicht versteuert. Zahl Deutschland 2022, vermutlich sogar die falsche Zahl.
  and existenzminimum:
    ‹einkommen ≤ 9888 ⟹ steuer einkommen = 0›


begin

end
```

Eigentlich hätte ich gerne noch eine weitere Anforderung. https://de.wikipedia.org/wiki/Steuerprogression sagt "Steuerprogression bedeutet das Ansteigen des Steuersatzes in Abhängigkeit vom zu versteuernden Einkommen oder Vermögen."

Formal betrachtet würde das bedeuten *einkommen-b* ≤ *einkommen-a* ⟹ (*λx. real-of-percentage* (*steuer-defs.steuersatz einkommen-b x*)) ≤ (*λx. real-of-percentage* (*steuer-defs.steuersatz einkommen-a x*))

Leider haben wir bereits jetzt in dem Modell eine Annahme getroffen, die es uns quasi unmöglich macht, ein Steuersystem zu implementieren, welches die Steuerprogression erfüllt. Der Grund ist, dass wir die Steuerfunktion auf ganzen Zahlen definiert haben. Aufgrund von Rundung können wir also immer Fälle haben, indem ein höheres Einkommen einen leicht geringeren Steuersatz hat als ein geringeres Einkommen. Beispielsweise bedeutet das für *beispiel-25prozent-steuer*, dass jemand mit 100 EUR Einkommen genau 25 Prozent Steuer zahlt, jemand mit 103 EUR Einkommen aber nur ca 24,3 Prozent Steuer zahlt.

```
lemma
  ‹steuer-defs.steuersatz beispiel-25prozent-steuer 100 = percentage 0.25›
  ‹steuer-defs.steuersatz beispiel-25prozent-steuer 103 = percentage (25 / 103)›
  ‹percentage (25 / 103) < percentage 0.25›
  ‹(103::nat) > 100›
```

In der Praxis sollten diese kleinen Rundungsfehler kein Problem darstellen, in diesem theoretischen Modell sorgen sie aber dafür, dass unser Steuersystem (und wir modellieren eine vereinfachte Version des deutschen Steuerystems) keine Steuerprogression erfüllt.

Die folgende Liste, basierend auf https://de.wikipedia.org/wiki/Einkommensteuer_(Deutschland)#Tarif_2022, sagt in welchem Bereich welcher Prozentsatz an Steuern zu zahlen ist. Beispielsweise sind die ersten *10347* steuerfrei.

```
definition steuerbuckets2022 :: ‹(nat × percentage) list› where
```

```
‹steuerbuckets2022 ≡ [
                (10347, percentage 0),
                (14926, percentage 0.14),
                (58596, percentage 0.2397),
                (277825, percentage 0.42)
               ]›
```

Für jedes Einkommen über 277825 gilt der Spitzensteuersatz von 45 Prozent. Wir ignorieren die Progressionsfaktoren in Zone 2 und 3.

Folgende Funktion berechnet die zu zahlende Steuer, basierend auf einer Steuerbucketliste.

```
fun bucketsteuerAbs :: ‹(nat × percentage) list ⇒ percentage ⇒ nat ⇒ real› where
  ‹bucketsteuerAbs ((bis, prozent)#mehr) spitzensteuer e =
      ((min bis e) * prozent)
      + (bucketsteuerAbs (map (λ(s,p). (s−bis,p)) mehr) spitzensteuer (e − bis))›
| ‹bucketsteuerAbs [] spitzensteuer e = e*spitzensteuer›
```

Die Einkommenssteuerberechnung, mit Spitzensteuersatz 45 Prozent und finalem Abrunden.

```
definition einkommenssteuer :: ‹nat ⇒ nat› where
 ‹einkommenssteuer einkommen ≡
   floor (bucketsteuerAbs steuerbuckets2022 (percentage 0.45) einkommen)›
```

Beispiel. Alles unter dem Existenzminimum ist steuerfrei:

```
lemma ‹einkommenssteuer 10 = 0›
lemma ‹einkommenssteuer 10000 = 0›
```

Für ein Einkommen nur knapp über dem Existenzminimum fällt sehr wenig Steuer an:

```
lemma ‹einkommenssteuer 14000 = floor ((14000−10347)*0.14)›
lemma ‹einkommenssteuer 14000 = 511›
```

Bei einem Einkommen von 20000 EUR wird ein Teil bereits mit den höheren Steuersatz der 3. Zone besteuert:

```
lemma ‹einkommenssteuer 20000 = 1857›
lemma ‹einkommenssteuer 20000 =
      floor ((14926−10347)*0.14 + (20000−14926)*0.2397)›
```

Höhere Einkommen führen zu einer höheren Steuer:

```
lemma ‹einkommenssteuer 40000 = 6651›
lemma ‹einkommenssteuer 60000 = 11698›
```

Die *einkommenssteuer* Funktion erfüllt die Anforderungen an *steuersystem*.

```
interpretation steuersystem
  where steuer = ‹einkommenssteuer›
```

19 Beispiel: Steuern

Wir nehmen eine einfach Welt an, in der jeder Person ihr Einkommen zugeordnet wird.

Achtung: Im Unterschied zum BeispielZahlenwelt.thy modellieren wir hier nicht den Gesamtbesitz, sondern das Jahreseinkommen. Besitz wird ignoriert.

datatype *steuerwelt = Steuerwelt*
    (*get-einkommen*: ‹*person ⇒ int*›) — einkommen jeder Person (im Zweifel 0).

fun *steuerwps* :: ‹*person ⇒ person ⇒ steuerwelt ⇒ steuerwelt*› **where**
  ‹*steuerwps p1 p2 (Steuerwelt besitz) = Steuerwelt (swap p1 p2 besitz)*›

fun *steuerlast* :: ‹*person ⇒ steuerwelt handlung ⇒ int*› **where**
  ‹*steuerlast p (Handlung vor nach) = ((get-einkommen vor) p) − ((get-einkommen nach) p)*›

fun *brutto* :: ‹*person ⇒ steuerwelt handlung ⇒ int*› **where**
  ‹*brutto p (Handlung vor nach) = (get-einkommen vor) p*›
fun *netto* :: ‹*person ⇒ steuerwelt handlung ⇒ int*› **where**
  ‹*netto p (Handlung vor nach) = (get-einkommen nach) p*›

lemma ‹*steuerlast Alice (Handlung (Steuerwelt 🌍[Alice:=8]) (Steuerwelt 🌍[Alice:=5])) = 3*›
lemma ‹*steuerlast Alice (Handlung (Steuerwelt 🌍[Alice:=8]) (Steuerwelt 🌍[Alice:=0])) = 8*›
lemma ‹*steuerlast Bob (Handlung (Steuerwelt 🌍[Alice:=8]) (Steuerwelt 🌍[Alice:=5])) = 0*›
lemma ‹*steuerlast Alice (Handlung (Steuerwelt 🌍[Alice:=−3]) (Steuerwelt 🌍[Alice:=−4])) = 1*›
lemma ‹*steuerlast Alice (Handlung (Steuerwelt 🌍[Alice:=1]) (Steuerwelt 🌍[Alice:=−1])) = 2*›

fun *mehrverdiener* :: ‹*person ⇒ steuerwelt handlung ⇒ person set*› **where**
  ‹*mehrverdiener ich (Handlung vor nach) = {p. (get-einkommen vor) p ≥ (get-einkommen vor) ich}*›

lemma ‹*mehrverdiener Alice*
    (Handlung (Steuerwelt 🌍[Alice:=8, Bob:=12, Eve:=7]) (Steuerwelt 🌍[Alice:=5]))
   = {Alice, Bob}›

lemma *mehrverdiener-betrachtet-nur-ausgangszustand*:
  ‹*mehrverdiener p (handeln p′ welt h) = mehrverdiener p (Handlung welt welt)*›

Folgende Maxime versucht Steuergerechtigkeit festzuschreiben:

definition *maxime-steuern* :: ‹*(person, steuerwelt) maxime*› **where**
  ‹*maxime-steuern ≡ Maxime*
   (λich handlung.
     (∀ p∈mehrverdiener ich handlung.
       steuerlast ich handlung ≤ steuerlast p handlung)
    ∧ (∀ p∈mehrverdiener ich handlung.
       netto ich handlung ≤ netto p handlung)

)›

thm *globale-maxime-katimp*

lemma ‹*wpsm-kommutiert* (*Maxime*
    (λ*ich handlung.*
        ($\forall$ *p*$\in$*mehrverdiener ich handlung.*
            steuerlast ich handlung $\leq$ *steuerlast p handlung*))) *steuerwps welt*›

lemma *wfh-steuerberechnung-jeder-zahlt-int*:
  ‹*ha* = *Handlungsabsicht* (λ*ich w. Steuerwelt* ((λ*e. e* − *steuerberechnung e*) $\circ$ (*get-einkommen w*)))
    $\Longrightarrow$ *wohlgeformte-handlungsabsicht steuerwps welt ha*›

thm *mehrverdiener-betrachtet-nur-ausgangszustand*

lemma ‹*ha* = *Handlungsabsicht* (λ*ich w. Steuerwelt* ((λ*e. e* − *steuerberechnung e*) $\circ$ (*get-einkommen w*)))
$\Longrightarrow$
  kategorischer-imperativ-auf ha welt
    (*Maxime*
      (λ*ich handlung.*
          ($\forall$ *p*$\in$*mehrverdiener ich handlung.*
              steuerlast ich handlung $\leq$ *steuerlast p handlung*)))›

TODO: finish, gilt aber nicht

Wenn die Steuerfunktion monoton ist, dann kann ich auch einen sehr eingeschraenken kat imp zeigen.

lemma ‹
  ($\bigwedge$*e1 e2. e1* $\leq$ *e2* $\Longrightarrow$ *steuerberechnung e1* $\leq$ *steuerberechnung e2*) $\Longrightarrow$
  ha = *Handlungsabsicht* (λ*ich w. Steuerwelt* ((λ*e. e* − *steuerberechnung e*) $\circ$ (*get-einkommen w*))) $\Longrightarrow$
  kategorischer-imperativ-auf ha welt
    (*Maxime*
      (λ*ich handlung.*
          ($\forall$ *p*$\in$*mehrverdiener ich handlung.*
              steuerlast ich handlung $\leq$ *steuerlast p handlung*)))›

19.1 Setup für Beispiele

```
definition ‹initialwelt ≡ Steuerwelt 🌍[Alice:=8, Bob:=3, Eve:= 5]›
```

```
definition ‹beispiel-case-law-absolut welt steuerfun ≡
  simulateOne
    (SimConsts
      Alice
      maxime-steuern
      (printable-case-law-ableiten-absolut (λw. show-fun (get-einkommen w))))
    3 steuerfun welt (Gesetz {})›
definition ‹beispiel-case-law-relativ welt steuerfun ≡
  simulateOne
    (SimConsts
      Alice
      maxime-steuern
      (case-law-ableiten-relativ delta-steuerwelt))
    1 steuerfun welt (Gesetz {})›
```

19.2 Beispiel: Keiner Zahlt Steuern

Die Maxime ist erfüllt, da wir immer nur kleiner-gleich fordern!

```
lemma ‹beispiel-case-law-relativ initialwelt (Handlungsabsicht (λich welt. welt)) =
  Gesetz {(§ 1, Rechtsnorm (Tatbestand []) (Rechtsfolge Erlaubnis))}›
```

19.3 Beispiel: Ich zahle 1 Steuer

Das funktioniert nicht:

```
definition ‹ich-zahle-1-steuer ich welt ≡
  Steuerwelt ((get-einkommen welt)(ich −= 1))›
lemma ‹beispiel-case-law-absolut initialwelt (Handlungsabsicht ich-zahle-1-steuer) =
  Gesetz
  {(§ 1,
    Rechtsnorm
     (Tatbestand
       ([(Alice, 8), (Bob, 3), (Carol, 0), (Eve, 5)],
        [(Alice, 7), (Bob, 3), (Carol, 0), (Eve, 5)]))
     (Rechtsfolge Verbot))}›
lemma ‹beispiel-case-law-relativ initialwelt (Handlungsabsicht ich-zahle-1-steuer) =
  Gesetz
  {(§ 1, Rechtsnorm (Tatbestand [Verliert Alice 1])
                    (Rechtsfolge Verbot))}›
```

Denn jeder muss Steuer zahlen! Ich finde es super spannend, dass hier faktisch ein Gleichbehandlungsgrundsatz rausfällt, ohne dass wir soewtas jemals explizit gefordert haben.

19.4 Beiepiel: Jeder zahle 1 Steuer

Jeder muss steuern zahlen: funktioniert, ist aber doof, denn am Ende sind alle im Minus.

Das *ich* wird garnicht verwendet, da jeder Steuern zahlt.

definition ‹*jeder-zahle-1-steuer ich welt* ≡
  Steuerwelt ((λ*e*. *e* − *1*) ∘ (*get-einkommen welt*))›
lemma ‹*beispiel-case-law-absolut initialwelt* (*Handlungsabsicht jeder-zahle-1-steuer*) =
Gesetz
  {(§ *3*,
    Rechtsnorm
     (*Tatbestand*
       ([(*Alice*, *6*), (*Bob*, *1*), (*Carol*, − *2*), (*Eve*, *3*)],
        [(*Alice*, *5*), (*Bob*, *0*), (*Carol*, − *3*), (*Eve*, *2*)]))
     (*Rechtsfolge Erlaubnis*)),
   (§ *2*,
    Rechtsnorm
     (*Tatbestand*
       ([(*Alice*, *7*), (*Bob*, *2*), (*Carol*, − *1*), (*Eve*, *4*)],
        [(*Alice*, *6*), (*Bob*, *1*), (*Carol*, − *2*), (*Eve*, *3*)]))
     (*Rechtsfolge Erlaubnis*)),
   (§ *1*,
    Rechtsnorm
     (*Tatbestand*
       ([(*Alice*, *8*), (*Bob*, *3*), (*Carol*, *0*), (*Eve*, *5*)],
        [(*Alice*, *7*), (*Bob*, *2*), (*Carol*, − *1*), (*Eve*, *4*)]))
     (*Rechtsfolge Erlaubnis*))}›
lemma ‹*beispiel-case-law-relativ initialwelt* (*Handlungsabsicht jeder-zahle-1-steuer*) =
  Gesetz
  {(§ *1*,
    Rechtsnorm
     (*Tatbestand* [*Verliert Alice 1*, *Verliert Bob 1*, *Verliert Carol 1*, *Verliert Eve 1*])
     (*Rechtsfolge Erlaubnis*))}›

19.5 Beispiel: Vereinfachtes Deutsches Steuersystem

Jetzt kommt die Steuern.thy ins Spiel.

definition *jeder-zahlt* :: ‹(*nat* ⇒ *nat*) ⇒ *'a* ⇒ *steuerwelt* ⇒ *steuerwelt*› **where**
  ‹*jeder-zahlt steuerberechnung ich welt* ≡
    Steuerwelt ((λ*e*. *e* − *steuerberechnung e*) ∘ *nat* ∘ (*get-einkommen welt*))›

definition ‹*jeder-zahlt-einkommenssteuer* ≡ *jeder-zahlt einkommenssteuer*›

Bei dem geringen Einkommen der *initialwelt* zahlt keiner Steuern.

lemma ‹*beispiel-case-law-absolut initialwelt* (*Handlungsabsicht jeder-zahlt-einkommenssteuer*) =
  Gesetz
  {(§ *1*,
    Rechtsnorm
     (*Tatbestand*
       ([(*Alice*, *8*), (*Bob*, *3*), (*Carol*, *0*), (*Eve*, *5*)],
        [(*Alice*, *8*), (*Bob*, *3*), (*Carol*, *0*), (*Eve*, *5*)]))
     (*Rechtsfolge Erlaubnis*))}›

Für höhere Einkommen erhalten wir plausible Werte und niemand rutscht ins negative:

lemma ‹*beispiel-case-law-relativ*
  (*Steuerwelt* 🌍[*Alice*:=*10000*, *Bob*:=*14000*, *Eve*:= *20000*])
  (*Handlungsabsicht jeder-zahlt-einkommenssteuer*)
  =
  Gesetz
  {(§ *1*,
    Rechtsnorm (*Tatbestand* [*Verliert Bob 511*, *Verliert Eve 1857*])
     (*Rechtsfolge Erlaubnis*))}›

20 Vereinfachtes Deutsches Steuersystem vs. die Steuermaxime

Die Anforderungen für ein *steuersystem* und die *maxime-steuern* sind vereinbar.

lemma *steuersystem-imp-maxime*:
  ‹*steuersystem steuersystem-impl* $\Longrightarrow$
     ($\forall$ *welt. moralisch welt maxime-steuern* (*Handlungsabsicht* (*jeder-zahlt steuersystem-impl*)))›

Danke ihr nats. Macht also keinen Sinn das als Annahme in die Maxime zu packen....

lemma *steuern-kleiner-einkommen-nat*:
    ‹*steuerlast ich* (*Handlung welt* (*jeder-zahlt steuersystem-impl ich welt*))
      $\leq$ *brutto ich* (*Handlung welt* (*jeder-zahlt steuersystem-impl ich welt*))›

lemma *maxime-imp-steuersystem*:
  ‹($\forall$ *einkommen. steuersystem-impl einkommen* $\leq$ *einkommen*) $\Longrightarrow$
    ($\forall$ *einkommen. einkommen* $\leq$ *9888* $\longrightarrow$ *steuersystem-impl einkommen* = *0*) $\Longrightarrow$
    $\forall$ *welt. moralisch welt maxime-steuern* (*Handlungsabsicht* (*jeder-zahlt steuersystem-impl*))
    $\Longrightarrow$ *steuersystem steuersystem-impl*›

Für jedes *steuersystem-impl*::*nat* $\Rightarrow$ *nat*, mit zwei weiteren Annahmen, gilt das *steuersystem* und *maxime-steuern* in der *jeder-zahlt* Implementierung äquivalent sind.

theorem
  fixes *steuersystem-impl* :: ‹*nat* $\Rightarrow$ *nat*›
  assumes *steuer-kleiner-einkommen*: ‹$\forall$ *einkommen. steuersystem-impl einkommen* $\leq$ *einkommen*›
    and *existenzminimum*: ‹$\forall$ *einkommen. einkommen* $\leq$ *9888* $\longrightarrow$ *steuersystem-impl einkommen* = *0*›
  shows
  ‹($\forall$ *welt. moralisch welt maxime-steuern* (*Handlungsabsicht* (*jeder-zahlt steuersystem-impl*)))
    $\longleftrightarrow$ *steuersystem steuersystem-impl*›