Stable Matching

Peter Gammie

March 19, 2025

Abstract

We mechanize proofs of several results from the *matching with contracts* literature, which generalize those of the classical two-sided matching scenarios that go by the name of *stable marriage*. Our focus is on game theoretic issues. Along the way we develop executable algorithms for computing optimal stable matches.

Contents

1 Introduction **2**

2 Sotomayor (1996): A non-constructive proof of the existence of stable marriages **2**

3 Preliminaries **7**
- 3.1 MaxR: maximum elements of linear orders . . . 7
- 3.2 Linear orders from lists . . . 8

4 Choice Functions **8**
- 4.1 The *substitutes* condition, AKA *independence of irrelevant alternatives* AKA *Chernoff* . . . 9
- 4.2 The *irrelevance of rejected contracts* condition AKA *consistency* AKA *Aizerman* . . . 11
- 4.3 The *law of aggregate demand* condition aka *size monotonicity* . . . 12
- 4.4 The *expansion* condition . . . 13
- 4.5 Axioms of revealed preference . . . 13
  - 4.5.1 The *strong axiom of revealed preference* ala Aygün and Sönmez (2012b) . . . 17
- 4.6 Choice functions arising from linear orders . . . 17
- 4.7 Plott's *path independence* condition . . . 19
  - 4.7.1 Path independence and decomposition into orderings . . . 19

5 Hatfield and Milgrom (2005): Matching with contracts **22**
- 5.1 Theorem 1: Existence of stable pairs . . . 24
  - 5.1.1 Theorem 1 does not hold (Aygün and Sönmez 2012b) . . . 25
  - 5.1.2 Theorem 1 holds with *independence of rejected contracts* . . . 26
  - 5.1.3 The converse of Theorem 1 . . . 27
- 5.2 Theorem 3: Algorithmics . . . 28
- 5.3 Theorem 4: Optimality . . . 31
- 5.4 Theorem 5 does not hold (Hatfield and Kojima 2008) . . . 32
- 5.5 Theorem 6: "Vacancy chain" dynamics . . . 33
- 5.6 Theorems 8 and 9: A "rural hospitals" theorem . . . 34
- 5.7 Theorems 15 and 16: Cumulative Offer Processes . . . 36
- 5.8 Concluding remarks . . . 39

6 Hatfield and Kojima (2010): Substitutes and stability for matching with contracts **39**
- 6.1 Theorem 1: the COP yields a stable match under *bilateral substitutes* . . . 40
- 6.2 Theorem 3: *pareto separability* relates *unilateral substitutes* and *substitutes* . . . 42

6.2.1 Afacan and Turhan (2015): *doctor separability* relates bi- and unilateral substitutes 44
6.3 Theorems 4 and 5: Doctor optimality 45
6.4 Theorem 6: A "rural hospitals" theorem 47
6.5 Concluding remarks 48

7 Kojima (2010): The non-existence of a stable and non-bossy mechanism 48

8 Strategic results 50
8.1 Hatfield and Milgrom (2005): Theorems 10 and 11: Truthful revelation as a Dominant Strategy . 50
8.2 Hatfield and Kojima (2009, 2010): The doctor-optimal match is group strategy-proof 53

9 Concluding remarks 55
9.1 Related work 55

10 Acknowledgements 56

References 59

1 Introduction

As economists have turned their attention to the design of such markets as school enrolments, internships, and housing refugees (Andersson and Ehlers 2016), particular *matching* scenarios have proven to be useful models. Roth (2015) defines matching as "economist-speak for how we get the many things we choose in life that also must choose us," and one such two-sided market is now colloquially known as the *stable marriage problem*. It was initially investigated by Gale and Shapley (1962), who introduced the key solution concept of *stability*, and the *deferred-acceptance algorithm* that efficiently constructs stable matches for it. We refer readers unfamiliar with this classical work to §2, where we formalize this scenario and mechanize a non-constructive existence proof of stable matches due to Sotomayor (1996). Further in-depth treatment can be found in the very readable monographs by Gusfield and Irving (1989) (algorithmics), Roth and Sotomayor (1990) (economics), and Manlove (2013).

Recently Hatfield and Milgrom (2005) (see also Fleiner (2000, 2002, 2003)) have recast the two-sided matching model to incorporate *contracts*, which intuitively allow agents to additionally indicate preferences over conditions such as salary. By allowing many-to-one matches, some aspects of a labour market can be modelled. Their analysis leans heavily on the lattice structure of the stable matches, and yields pleasingly simple and general algorithms (§5). Later work trades this structure for generality, and the analysis becomes more intricate (§6). The key game-theoretic result is the (one-sided) strategy-proofness of the optimal stable match (§8).

This work was motivated by the difficulty of navigating the literature on *matching with contracts* by non-specialists, as observed by Caminati et al. (2015a,b). We impose some order by formalizing much of it in Isabelle/HOL (Nipkow et al. 2002), a proof assistant for a simply-typed higher-order logic. By carefully writing definitions that are executable and testable, we avail ourselves of Isabelle's automatic tools, specifically `nitpick` and `sledgehammer`, to rapidly identify errors when formulating assertions. We focus primarily on strategic (game theoretic) issues, but our development is also intended to serve as a foundation for further results.

The proof assistant forces us to take care of all details, which yields a verbosity that may deter some readers. We suggest that most will fare best by reading the definitions and **lemma/theorem** statements closely, and skipping the proofs. (The important results are labelled **theorem** and **proposition**, but often the **lemma**s contain the meat.) The material in §4 on choice functions is mostly for reference.

This PDF is generated directly from the development's sources and is extensively hyperlinked, but for some purposes there is no substitute to firing up Isabelle.

2 Sotomayor (1996): A non-constructive proof of the existence of stable marriages

We set the scene with a non-constructive proof of the existence of stable matches due to Sotomayor (1996). This approach is pleasantly agnostic about the strictness of preferences, and moreover avoids getting bogged down in reasoning about programs; most existing proofs involve such but omit formal treatments of the requisite assertions. This tradition started with Gale and Shapley (1962); see Bijlsma (1991) for a rigorous treatment.

The following contains the full details of an Isabelle/HOL formalization of her proof, and aims to introduce the machinery we will make heavy use of later. Further developments will elide many of the more tedious technicalities that we include here.

The scenario consists of disjoint finite sets of men M and women W, represented as types $'m{::}\mathit{finite}$ and $'w{::}\mathit{finite}$ respectively. We diverge from Sotomayor by having each man and woman rank only acceptable partners in a way that is transitive and complete. (Here completeness requires *Refl* in addition to *Total* as the latter does not imply the former, and so we end up with a total preorder.) Such orders therefore include cycles of indifference, i.e., are not antisymmetric.

Also matches are treated as relations rather than functions.

We model this scenario in a **locale**, a sectioning mechanism for stating a series of lemmas relative to a set of fixed variables (**fixes**) and assumptions (**assumes**) that can later be instantiated and discharged.

type-synonym $('m, 'w)\ \mathit{match} = ('m \times 'w)\ \mathit{set}$

locale *StableMarriage* =
  fixes $\mathit{Pm} :: 'm{::}\mathit{finite} \Rightarrow 'w{::}\mathit{finite}\ \mathit{rel}$
  fixes $\mathit{Pw} :: 'w \Rightarrow 'm\ \mathit{rel}$
  assumes $\mathit{Pm\text{-}pref}$: $\forall m.\ \mathit{Preorder}\ (\mathit{Pm}\ m) \wedge \mathit{Total}\ (\mathit{Pm}\ m)$
  assumes $\mathit{Pw\text{-}pref}$: $\forall w.\ \mathit{Preorder}\ (\mathit{Pw}\ w) \wedge \mathit{Total}\ (\mathit{Pw}\ w)$
begin

A *match* assigns at most one man to each woman, and vice-versa. It is also *individually rational*, i.e., the partners are acceptable to each other. The constant *Field* is the union of the *Domain* and *Range* of a relation.

definition $\mathit{match} :: ('m, 'w)\ \mathit{match} \Rightarrow \mathit{bool}$ **where**
  $\mathit{match}\ \mu \longleftrightarrow \mathit{inj\text{-}on}\ \mathit{fst}\ \mu \wedge \mathit{inj\text{-}on}\ \mathit{snd}\ \mu \wedge \mu \subseteq (\bigcup m.\ \{m\} \times \mathit{Field}\ (\mathit{Pm}\ m)) \cap (\bigcup w.\ \mathit{Field}\ (\mathit{Pw}\ w) \times \{w\})$

A woman *prefers* one man to another if her preference order ranks the former over the latter, and *strictly prefers* him if additionally the latter is not ranked over the former, and similarly for the men.

abbreviation (*input*) $\mathit{m\text{-}for}\ w\ \mu \equiv \{m.\ (m, w) \in \mu\}$
abbreviation (*input*) $\mathit{w\text{-}for}\ m\ \mu \equiv \{w.\ (m, w) \in \mu\}$

definition $\mathit{m\text{-}prefers} :: 'm \Rightarrow ('m, 'w)\ \mathit{match} \Rightarrow 'w\ \mathit{set}$ **where**
  $\mathit{m\text{-}prefers}\ m\ \mu = \{w' \in \mathit{Field}\ (\mathit{Pm}\ m).\ \forall w{\in}\mathit{w\text{-}for}\ m\ \mu.\ (w, w') \in \mathit{Pm}\ m\}$

definition $\mathit{w\text{-}prefers} :: 'w \Rightarrow ('m, 'w)\ \mathit{match} \Rightarrow 'm\ \mathit{set}$ **where**
  $\mathit{w\text{-}prefers}\ w\ \mu = \{m' \in \mathit{Field}\ (\mathit{Pw}\ w).\ \forall m{\in}\mathit{m\text{-}for}\ w\ \mu.\ (m, m') \in \mathit{Pw}\ w\}$

definition $\mathit{m\text{-}strictly\text{-}prefers} :: 'm \Rightarrow ('m, 'w)\ \mathit{match} \Rightarrow 'w\ \mathit{set}$ **where**
  $\mathit{m\text{-}strictly\text{-}prefers}\ m\ \mu = \{w' \in \mathit{Field}\ (\mathit{Pm}\ m).\ \forall w{\in}\mathit{w\text{-}for}\ m\ \mu.\ (w, w') \in \mathit{Pm}\ m \wedge (w', w) \notin \mathit{Pm}\ m\}$

definition $\mathit{w\text{-}strictly\text{-}prefers} :: 'w \Rightarrow ('m, 'w)\ \mathit{match} \Rightarrow 'm\ \mathit{set}$ **where**
  $\mathit{w\text{-}strictly\text{-}prefers}\ w\ \mu = \{m' \in \mathit{Field}\ (\mathit{Pw}\ w).\ \forall m{\in}\mathit{m\text{-}for}\ w\ \mu.\ (m, m') \in \mathit{Pw}\ w \wedge (m', m) \notin \mathit{Pw}\ w\}$

A couple *blocks* a match μ if both strictly prefer each other to anyone they are matched with in μ.

definition $\mathit{blocks} :: 'm \Rightarrow 'w \Rightarrow ('m, 'w)\ \mathit{match} \Rightarrow \mathit{bool}$ **where**
  $\mathit{blocks}\ m\ w\ \mu \longleftrightarrow w \in \mathit{m\text{-}strictly\text{-}prefers}\ m\ \mu \wedge m \in \mathit{w\text{-}strictly\text{-}prefers}\ w\ \mu$

We say a match is *stable* if there are no blocking couples.

definition $\mathit{stable} :: ('m, 'w)\ \mathit{match} \Rightarrow \mathit{bool}$ **where**
  $\mathit{stable}\ \mu \longleftrightarrow \mathit{match}\ \mu \wedge (\forall m\ w.\ \neg\ \mathit{blocks}\ m\ w\ \mu)$

lemma *stable-match*:
  assumes $\mathit{stable}\ \mu$
  shows $\mathit{match}\ \mu$
⟨*proof*⟩

Our goal is to show that for every preference order there is a stable match. Stable matches in this scenario form a lattice, and this proof implicitly adopts the traditional view that men propose and women choose.

The definitions above form the trust basis for this existence theorem; the following are merely part of the proof apparatus, and Isabelle/HOL enforces their soundness with respect to the argument. We will see these concepts again in later developments.

Firstly, a match is *simple* if every woman party to a blocking pair is single. The most obvious such match leaves everyone single.

definition *simple* :: $('m, 'w)$ *match* $\Rightarrow$ *bool* **where**
  simple $\mu \longleftrightarrow$ *match* $\mu \wedge (\forall\, m\ w.$ *blocks* $m\ w\ \mu \longrightarrow w \notin$ *Range* $\mu)$

lemma *simple-match*:
  assumes *simple* μ
  shows *match* μ
⟨*proof*⟩

lemma *simple-ex*:
  $\exists\, \mu.$ *simple* μ
⟨*proof*⟩

Sotomayor observes the following:

lemma *simple-no-single-women-stable*:
  assumes *simple* μ
  assumes $\forall\, w.\ w \in$ *Range* μ — No woman is single
  shows *stable* μ
⟨*proof*⟩

lemma *stable-simple*:
  assumes *stable* μ
  shows *simple* μ
⟨*proof*⟩

Secondly, a *weakly Pareto optimal match for men (among all simple matches)* is one for which there is no other match that all men like as much and some man likes more.

definition *m-weakly-prefers* :: $'m \Rightarrow ('m, 'w)$ *match* $\Rightarrow$ $'w$ *set* **where**
  m-weakly-prefers $m\ \mu = \{w' \in$ *Field* $($*Pm* $m).\ \forall\, w{\in}$*w-for* $m\ \mu.\ (w, w') \in$ *Pm* $m\}$

definition *weakly-preferred-by-men* :: $('m, 'w)$ *match* $\Rightarrow ('m, 'w)$ *match* $\Rightarrow$ *bool* **where**
  weakly-preferred-by-men $\mu\ \mu'$
    $\longleftrightarrow (\forall\, m.\ \forall\, w{\in}$*w-for* $m\ \mu.\ \exists\, w'{\in}$*w-for* $m\ \mu'.\ w' \in$ *m-weakly-prefers* $m\ \mu)$

definition *strictly-preferred-by-a-man* :: $('m, 'w)$ *match* $\Rightarrow ('m, 'w)$ *match* $\Rightarrow$ *bool* **where**
  strictly-preferred-by-a-man $\mu\ \mu'$
    $\longleftrightarrow (\exists\, m.\ \exists\, w{\in}$*w-for* $m\ \mu'.\ w \in$ *m-strictly-prefers* $m\ \mu)$

definition *weakly-Pareto-optimal-for-men* :: $('m, 'w)$ *match* $\Rightarrow$ *bool* **where**
  weakly-Pareto-optimal-for-men μ
    $\longleftrightarrow$ *simple* $\mu \wedge \neg(\exists\, \mu'.$ *simple* $\mu' \wedge$ *weakly-preferred-by-men* $\mu\ \mu' \wedge$ *strictly-preferred-by-a-man* $\mu\ \mu')$

We will often provide *introduction rules* for more complex predicates, and sometimes derive these by elementary syntactic manipulations expressed by the *attributes* enclosed in square brackets after a use-mention of a lemma. The **lemmas** command binds a name to the result. To conform with the Isar structured proof language, we use meta-logic ("Pure" in Isabelle terminology) connectives: $\bigwedge$ denotes universal quantification, and $\Longrightarrow$ implication.

lemma *weakly-preferred-by-menI*:
  assumes $\bigwedge m\ w.\ (m, w) \in \mu \Longrightarrow \exists\, w'.\ (m, w') \in \mu' \wedge w' \in$ *m-weakly-prefers* $m\ \mu$
  shows *weakly-preferred-by-men* $\mu\ \mu'$
⟨*proof*⟩

lemmas *simpleI* = *iffD2*[*OF simple-def*, *unfolded conj-imp-eq-imp-imp*, *rule-format*]

lemma *weakly-Pareto-optimal-for-men-simple*:
  assumes *weakly-Pareto-optimal-for-men* μ
  shows *simple* μ
$\langle proof \rangle$

Later we will elide obvious technical lemmas like the following. The more obscure proofs are typically generated automatically by sledgehammer (Blanchette et al. 2016).

lemma *m-weakly-prefers-Pm*:
  assumes *match* μ
  assumes $(m, w) \in \mu$
  shows $w' \in$ *m-weakly-prefers* $m\ \mu \longleftrightarrow (w, w') \in Pm\ m$
$\langle proof \rangle$

lemma *match-Field*:
  assumes *match* μ
  assumes $(m, w) \in \mu$
  shows $w \in Field\ (Pm\ m)$
    and $m \in Field\ (Pw\ w)$
$\langle proof \rangle$

lemma *weakly-preferred-by-men-refl*:
  assumes *match* μ
  shows *weakly-preferred-by-men* $\mu\ \mu$
$\langle proof \rangle$

Sotomayor, p137 provides an alternative definition of *weakly-preferred-by-men*. The syntax (**is** *?lhs* $\longleftrightarrow$ *pat*) binds the *schematic variables ?lhs* and *?rhs* to the terms separated by $\longleftrightarrow$.

lemma *weakly-preferred-by-men-strictly-preferred-by-a-man*:
  assumes *match* μ
  assumes *match* μ'
  shows *weakly-preferred-by-men* $\mu\ \mu' \longleftrightarrow \neg$*strictly-preferred-by-a-man* $\mu'\ \mu$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)
$\langle proof \rangle$

lemma *weakly-Pareto-optimal-for-men-def2*:
  weakly-Pareto-optimal-for-men μ
    $\longleftrightarrow$ *simple* $\mu \wedge (\forall \mu'.$ *simple* $\mu' \wedge$ *strictly-preferred-by-a-man* $\mu\ \mu' \longrightarrow$ *strictly-preferred-by-a-man* $\mu'\ \mu)$
$\langle proof \rangle$

Sotomayor claims that the existence of such a weakly Pareto optimal match for men is "guaranteed by the fact that *the set of simple matchings is nonempty* [our *simple-ex* lemma] *and finite and the preferences are transitive.*" The following lemmas express this intuition:

lemma *trans-finite-has-maximal-elt*:
  assumes *trans* r
  assumes *finite* $(Field\ r)$
  assumes $Field\ r \neq \{\}$
  shows $\exists x \in Field\ r.\ (\forall y \in Field\ r.\ (x, y) \in r \longrightarrow (y, x) \in r)$
$\langle proof \rangle$

lemma *weakly-Pareto-optimal-for-men-ex*:
  $\exists \mu.$ *weakly-Pareto-optimal-for-men* μ
$\langle proof \rangle$

The main result proceeds by contradiction.

lemma *weakly-Pareto-optimal-for-men-stable*:
  assumes *weakly-Pareto-optimal-for-men* μ
  shows *stable* μ
$\langle proof \rangle$

theorem *stable-ex*:
  $\exists \mu.\ stable\ \mu$
$\langle proof \rangle$

We can exit the locale context and later re-enter it.

end

We *interpret* the locale by supplying constants that instantiate the variables we fixed earlier, and proving that these satisfy the assumptions. In this case we provide concrete preference orders, and by doing so we demonstrate that our theory is non-vacuous. We arbitrarily choose Roth and Sotomayor (1990, Example 2.15) which demonstrates the non-existence of man- or woman-optimal matches if preferences are non-strict. (We define optimality shortly.) The following bunch of types eases the description of this particular scenario.

datatype $M = M1 \mid M2 \mid M3$
datatype $W = W1 \mid W2 \mid W3$

lemma *M-UNIV*: $UNIV = set\ [M1,\ M2,\ M3]$ $\langle proof \rangle$
lemma *W-UNIV*: $UNIV = set\ [W1,\ W2,\ W3]$ $\langle proof \rangle$

instance M :: *finite* $\langle proof \rangle$
instance W :: *finite* $\langle proof \rangle$

lemma *M-All*:
  shows $(\forall m.\ P\ m) \longleftrightarrow (\forall m \in set\ [M1,\ M2,\ M3].\ P\ m)$
$\langle proof \rangle$

lemma *W-All*:
  shows $(\forall w.\ P\ w) \longleftrightarrow (\forall w \in set\ [W1,\ W2,\ W3].\ P\ w)$
$\langle proof \rangle$

primrec $Pm :: M \Rightarrow W\ rel$ **where**
  $Pm\ M1 = \{\ (W1,\ W1), (W1,\ W2), (W1,\ W3), (W2,\ W2), (W2,\ W3), (W3,\ W3), (W3,\ W2)\ \}$
$\mid Pm\ M2 = \{\ (W1,\ W1), (W1,\ W2), (W2,\ W2)\ \}$
$\mid Pm\ M3 = \{\ (W1,\ W1), (W1,\ W3), (W3,\ W3)\ \}$

primrec $Pw :: W \Rightarrow M\ rel$ **where**
  $Pw\ W1 = \{\ (M3,\ M3), (M3,\ M2), (M3,\ M1), (M2,\ M2), (M2,\ M1), (M1,\ M1)\ \}$
$\mid Pw\ W2 = \{\ (M2,\ M2), (M2,\ M1), (M1,\ M1)\ \}$
$\mid Pw\ W3 = \{\ (M3,\ M3), (M3,\ M1), (M1,\ M1)\ \}$

lemma *Pm*: $Preorder\ (Pm\ m) \wedge Total\ (Pm\ m)$
$\langle proof \rangle$

lemma *Pw*: $Preorder\ (Pw\ w) \wedge Total\ (Pw\ w)$
$\langle proof \rangle$

interpretation *Non-Strict*: *StableMarriage Pm Pw*
$\langle proof \rangle$

We demonstrate that there are only two stable matches in this scenario. Isabelle/HOL does not have any special support for these types of model checking problems, so we simply try all combinations of men and women. Clearly this does not scale, and for larger domains we need to be a bit cleverer (see §7).

lemma *Non-Strict-stable1*:
  shows *Non-Strict.stable* $\{(M1,\ W2), (M2,\ W1), (M3,\ W3)\}$
$\langle proof \rangle$

lemma *Non-Strict-stable2*:
  shows *Non-Strict.stable* $\{(M1,\ W3), (M2,\ W2), (M3,\ W1)\}$
$\langle proof \rangle$

lemma *Non-Strict-stable-matches*:
  Non-Strict.stable μ
    $\longleftrightarrow \mu = \{(M1,\ W2),\ (M2,\ W1),\ (M3,\ W3)\}$
    $\vee\ \mu = \{(M1,\ W3),\ (M2,\ W2),\ (M3,\ W1)\}$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)
$\langle proof \rangle$

So far the only interesting result in this interpretation of *StableMarriage* is the *Non-Strict.stable-ex* theorem, i.e., that there is a stable match. We now add the notion of *optimality* to our locale, and all interpretations will automatically inherit it. Later we will also extend locales by adding new fixed variables and assumptions.

Following Roth and Sotomayor (1990, Definition 2.11), a stable match is *optimal for men* if every man likes it at least as much as any other stable match (and similarly for an *optimal for women* match).

context *StableMarriage*
begin

definition *optimal-for-men* :: $('m,\ 'w)$ *match* $\Rightarrow$ *bool* **where**
  optimal-for-men μ
    $\longleftrightarrow$ *stable* $\mu \wedge (\forall\, \mu'.$ *stable* $\mu' \longrightarrow$ *weakly-preferred-by-men* $\mu'\ \mu)$

definition *w-weakly-prefers* :: $'w \Rightarrow ('m,\ 'w)$ *match* $\Rightarrow\ 'm$ *set* **where**
  w-weakly-prefers $w\ \mu = \{m' \in Field\ (Pw\ w).\ \forall\, m{\in}m\text{-}for\ w\ \mu.\ (m,\ m') \in Pw\ w\}$

definition *weakly-preferred-by-women* :: $('m,\ 'w)$ *match* $\Rightarrow ('m,\ 'w)$ *match* $\Rightarrow$ *bool* **where**
  weakly-preferred-by-women $\mu\ \mu'$
    $\longleftrightarrow (\forall\, w.\ \forall\, m{\in}m\text{-}for\ w\ \mu.\ \exists\, m'{\in}m\text{-}for\ w\ \mu'.\ m' \in w\text{-}weakly\text{-}prefers\ w\ \mu)$

definition *optimal-for-women* :: $('m,\ 'w)$ *match* $\Rightarrow$ *bool* **where**
  optimal-for-women μ
    $\longleftrightarrow$ *stable* $\mu \wedge (\forall\, \mu'.$ *stable* $\mu' \longrightarrow$ *weakly-preferred-by-women* $\mu\ \mu')$

end

We can show that there is no optimal stable match for these preferences:

lemma *NonStrict-not-optimal*:
  assumes *Non-Strict.stable* μ
  shows $\neg$*Non-Strict.optimal-for-men* $\mu \wedge \neg$*Non-Strict.optimal-for-women* μ
$\langle proof \rangle$

Sotomayor (1996) remarks that, if the preferences are strict, there is only one weakly Pareto optimal match for men, and that it is man-optimal. (This is the match found by the classic man-proposing deferred acceptance algorithm due to Gale and Shapley (1962).) However she omits a proof that the man-optimal match actually exists under strict preferences.

The easiest way to show this and further results is to exhibit the lattice structure of the stable matches discovered by Conway (see Roth and Sotomayor (1990, Theorem 2.16)), where the men- and women-optimal matches are the extremal points. This suggests looking for a monotonic function whose fixed points are this lattice, which is the essence of the analysis of matching with contracts in §5.

3 Preliminaries

$\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle\langle proof \rangle$

3.1 MaxR: maximum elements of linear orders

We generalize the existing *max* and *Max* functions to work on orders defined over sets. See §4.6 for choice-function related lemmas.

locale *MaxR* =
  fixes r :: $'a$::*finite rel*

assumes *r-Linear-order*: *Linear-order r*
begin

The basic function chooses the largest of two elements:

definition *maxR* :: $'a \Rightarrow 'a \Rightarrow 'a$ **where**
maxR x y = (*if* (*x*, *y*) $\in$ *r then y else x*)⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

We hoist this to finite sets using the *Finite-Set.fold* combinator. For code generation purposes it seems inevitable that we need to fuse the fold and filter into a single total recursive definition.

definition *MaxR-f* :: $'a \Rightarrow 'a$ *option* $\Rightarrow 'a$ *option* **where**
MaxR-f x acc = (*if x* $\in$ *Field r then Some* (*case acc of None* $\Rightarrow$ *x* | *Some y* $\Rightarrow$ *maxR x y*) *else acc*)

interpretation *MaxR-f*: *comp-fun-idem MaxR-f*

definition *MaxR-opt* :: $'a$ *set* $\Rightarrow 'a$ *option* **where**
MaxR-opt-eq-fold': *MaxR-opt A* = *Finite-Set.fold MaxR-f None A*⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*pro
end

interpretation *MaxR-empty*: *MaxR* {}
⟨*proof*⟩

interpretation *MaxR-singleton*: *MaxR* {(*x*,*x*)} **for** *x*
⟨*proof*⟩

lemma *MaxR-r-domain* [*iff*]:
assumes *MaxR r*
shows *MaxR* (*Restr r A*)
⟨*proof*⟩

3.2 Linear orders from lists

Often the easiest way to specify a concrete linear order is with a list. Here these run from greatest to least.

primrec *linord-of-listP* :: $'a \Rightarrow 'a \Rightarrow 'a$ *list* $\Rightarrow$ *bool* **where**
linord-of-listP x y [] $\longleftrightarrow$ *False*
| *linord-of-listP x y* (*z* # *zs*) $\longleftrightarrow$ (*z* = *y* $\wedge$ *x* $\in$ *set* (*z* # *zs*)) $\vee$ *linord-of-listP x y zs*

definition *linord-of-list* :: $'a$ *list* $\Rightarrow 'a$ *rel* **where**
linord-of-list xs $\equiv$ {(*x*, *y*). *linord-of-listP x y xs*}
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

lemma *linord-of-list-Linear-order*:
assumes *distinct xs*
assumes *ys* = *set xs*
shows *linear-order-on ys* (*linord-of-list xs*)

Every finite linear order is generated by a list.

⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*pro
lemma *linear-order-on-list*:
assumes *linear-order-on ys r*
assumes *ys* = *Field r*
assumes *finite ys*
shows $\exists!$*xs*. *r* = *linord-of-list xs* $\wedge$ *distinct xs* $\wedge$ *set xs* = *ys*

4 Choice Functions

We now develop a few somewhat general results about choice functions, following Border (2012); Moulin (1985); Sen (1970). Hansson and Grüne-Yanoff (2012) provide some philosophical background on this topic. While this material is foundational to the story we tell about stable matching, it is perhaps best skipped over on a first reading.

The game here is to study conditions on functions that yield acceptable choices from a given set of alternatives drawn from some universe (a set, often a type in HOL). We adopt the Isabelle convention of attaching the suffix *on* to predicates that are defined on subsets of their types.

type-synonym $'a\ cfun = \ 'a\ set \Rightarrow \ 'a\ set$

Most results require that the choice function yield a subset of its argument:

definition *f-range-on* :: $'a\ set \Rightarrow \ 'a\ cfun \Rightarrow bool$ **where**
  f-range-on $A\ f \longleftrightarrow (\forall B \subseteq A.\ f\ B \subseteq B)$

abbreviation *f-range* :: $'a\ cfun \Rightarrow bool$ **where**
  f-range $\equiv$ *f-range-on* $UNIV \langle proof \rangle \langle proof \rangle$

Economists typically assume that the universe is finite, and f is *decisive*, i.e., yields non-empty sets when given non-empty sets.

definition *decisive-on* :: $'a\ set \Rightarrow \ 'a\ cfun \Rightarrow bool$ **where**
  decisive-on $A\ f \longleftrightarrow (\forall B \subseteq A.\ B \neq \{\} \longrightarrow f\ B \neq \{\})$

abbreviation *decisive* :: $'a\ cfun \Rightarrow bool$ **where**
  decisive $\equiv$ *decisive-on* $UNIV \langle proof \rangle \langle proof \rangle$

Often we can mildly generalise existing results by not requiring that f be *decisive*, and by dropping the finiteness hypothesis. We make essential use of the former generalization in §5.

Some choice functions, such as those arising from linear orders (§4.6), are *resolute*: these always yield a single choice.

definition *resolute-on* :: $'a\ set \Rightarrow \ 'a\ cfun \Rightarrow bool$ **where**
  resolute-on $A\ f \longleftrightarrow (\forall B \subseteq A.\ B \neq \{\} \longrightarrow (\exists a.\ f\ B = \{a\}))$

abbreviation *resolute* :: $'a\ cfun \Rightarrow bool$ **where**
  resolute $\equiv$ *resolute-on* $UNIV$

lemma *resolute-on-decisive-on*:
  assumes *resolute-on* $A\ f$
  shows *decisive-on* $A\ f$

Often we talk about the choices that are rejected by f:

abbreviation *Rf* :: $'a\ cfun \Rightarrow \ 'a\ cfun$ **where**
  $Rf\ f\ X \equiv X - f\ X$

Typically there are many (almost-)equivalent formulations of each property in the literature. We try to formulate our rules in terms of the most general of these.

4.1 The *substitutes* condition, AKA *independence of irrelevant alternatives* AKA *Chernoff*

Loosely speaking, the *substitutes* condition asserts that an alternative that is rejected from A shall remain rejected when there is "increased competition," i.e., from all sets that contain A.

Hatfield and Milgrom (2005) define this property as simply the monotonicity of *Rf*. Aygün and Sönmez (2012b) instead use the complicated condition shown here. Condition α, due to Sen (1970, p17, see below), is the most general and arguably the most perspicuous.

definition *substitutes-on* :: $'a\ set \Rightarrow \ 'a\ cfun \Rightarrow bool$ **where**
  substitutes-on $A\ f \longleftrightarrow \neg(\exists B \subseteq A.\ \exists a\ b.\ \{a, b\} \subseteq A - B \wedge b \notin f\ (B \cup \{b\}) \wedge b \in f\ (B \cup \{a, b\}))$

abbreviation *substitutes* :: $'a$ *cfun* $\Rightarrow$ *bool* **where**
substitutes $\equiv$ *substitutes-on UNIV*

lemma *substitutes-on-def2*[*simplified*]:
substitutes-on $A\ f \longleftrightarrow (\forall\, B{\subseteq}A.\ \forall\, a{\in}A.\ \forall\, b{\in}A.\ b \notin f\ (B \cup \{b\}) \longrightarrow b \notin f\ (B \cup \{a,\ b\}))$⟨*proof*⟩

lemma *substitutes-on-union*:
assumes $a \notin f\ (B \cup \{a\})$
assumes *substitutes-on* $(A \cup B \cup \{a\})\ f$
assumes *finite* A
shows $a \notin f\ (A \cup B \cup \{a\})$

lemma *substitutes-on-antimono*:
assumes *substitutes-on* $B\ f$
assumes $A \subseteq B$
shows *substitutes-on* $A\ f$

The equivalence with the monotonicity of alternative-rejection requires a finiteness constraint.

lemma *substitutes-on-Rf-mono-on*:
assumes *substitutes-on* $A\ f$
assumes *finite* A
shows *mono-on* (*Pow* A) (*Rf* f)

lemma *Rf-mono-on-substitutes*:
assumes *mono-on* (*Pow* A) (*Rf* f)
shows *substitutes-on* $A\ f$

The above substitutes condition is equivalent to the *independence of irrelevant alternatives*, AKA condition α due to Sen (1970). Intuitively if a is chosen from a set A, then it must be chosen from every subset of A that it belongs to. Note the lack of finiteness assumptions here.

definition *iia-on* :: $'a$ *set* $\Rightarrow$ $'a$ *cfun* $\Rightarrow$ *bool* **where**
iia-on $A\ f \longleftrightarrow (\forall\, B{\subseteq}A.\ \forall\, C{\subseteq}B.\ \forall\, a{\in}C.\ a \in f\ B \longrightarrow a \in f\ C)$

abbreviation *iia* :: $'a$ *cfun* $\Rightarrow$ *bool* **where**
iia $\equiv$ *iia-on UNIV*

lemma *Rf-mono-on-iia-on*:
shows *mono-on* (*Pow* A) (*Rf* f) $\longleftrightarrow$ *iia-on* $A\ f$

lemma *Rf-mono-iia*:
shows *mono* (*Rf* f) $\longleftrightarrow$ *iia* f

lemma *substitutes-iia*:
assumes *finite* A
shows *substitutes-on* $A\ f$ $\longleftrightarrow$ *iia-on* $A\ f$

One key result is that the choice function must be idempotent if it satisfies *iia* or any of the equivalent conditions.

lemma *iia-f-idem*:
assumes *f-range-on* $A\ f$
assumes *iia-on* $A\ f$
assumes $B \subseteq A$
shows $f\ (f\ B) = f\ B$

Hatfield and Milgrom (2005, p914, bottom right) claim that the *substitutes* condition coincides with the *substitutable preferences* condition for the college admissions problem of Roth and Sotomayor (1990, Definition 6.2), which is similar to *iia*:

definition *substitutable-preferences-on* :: $'a$ *set* $\Rightarrow$ $'a$ *cfun* $\Rightarrow$ *bool* **where**
substitutable-preferences-on $A\ f \longleftrightarrow (\forall B{\subseteq}A.\ \forall a{\in}B.\ \forall b{\in}B.\ a \neq b \wedge a \in f\ B \longrightarrow a \in f\ (B - \{b\}))$

lemma *substitutable-preferences-on-substitutes-on*:
shows *substitutable-preferences-on* $A\ f \longleftrightarrow$ *substitutes-on* $A\ f$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)

Moulin (1985, p152) defines an equivalent *Chernoff* condition. Intuitively this captures the idea that "a best choice in some issue [set of alternatives] is still best if the issue shrinks."

definition *Chernoff-on* :: $'a$ *set* $\Rightarrow$ $'a$ *cfun* $\Rightarrow$ *bool* **where**
Chernoff-on $A\ f \longleftrightarrow (\forall B{\subseteq}A.\ \forall C{\subseteq}B.\ f\ B \cap C \subseteq f\ C)$

abbreviation *Chernoff* :: $'a$ *cfun* $\Rightarrow$ *bool* **where**
Chernoff $\equiv$ *Chernoff-on UNIV*

lemmas *Chernoff-onI* = *iffD2*[*OF Chernoff-on-def*, *rule-format*]
lemmas *Chernoff-def* = *Chernoff-on-def*[**where** *A*=*UNIV*, *simplified*]

lemma *Chernoff-on-iia-on*:
shows *Chernoff-on* $A\ f \longleftrightarrow$ *iia-on* $A\ f$

lemma *Chernoff-on-union*:
assumes *Chernoff-on* $A\ f$
assumes *f-range-on* $A\ f$
assumes $B \subseteq A\ \ C \subseteq A$
shows $f\ (B \cup C) \subseteq f\ B \cup f\ C$

Moulin (1985, p159) states a series of equivalent formulations of the *Chernoff* condition. He also claims that these hold if the two sets are disjoint.

lemma *Chernoff-a*:
assumes *f-range-on* $A\ f$
shows *Chernoff-on* $A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) \subseteq f\ B \cup C)$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)

lemma *Chernoff-b*: — essentially the converse of *Chernoff-on-union*
assumes *f-range-on* $A\ f$
shows *Chernoff-on* $A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) \subseteq f\ B \cup f\ C)$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)

lemma *Chernoff-c*:
assumes *f-range-on* $A\ f$
shows *Chernoff-on* $A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) \subseteq f\ (f\ B \cup C))$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)

lemma *Chernoff-d*:
assumes *f-range-on* $A\ f$
shows *Chernoff-on* $A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) \subseteq f\ (f\ B \cup f\ C))$ (**is** *?lhs* $\longleftrightarrow$ *?rhs*)

4.2 The *irrelevance of rejected contracts* condition AKA *consistency* AKA *Aizerman*

Aygün and Sönmez (2012b, §4) propose to repair the results of Hatfield and Milgrom (2005) by imposing the *irrelevance of rejected contracts* (IRC) condition. Intuitively this requires the choice function f to ignore unchosen alternatives.

definition *irc-on* :: $'a$ *set* $\Rightarrow$ $'a$ *cfun* $\Rightarrow$ *bool* **where**
irc-on $A\ f \longleftrightarrow (\forall B{\subseteq}A.\ \forall a{\in}A.\ a \notin f\ (B \cup \{a\}) \longrightarrow f\ (B \cup \{a\}) = f\ B)$

abbreviation *irc* :: *'a cfun* ⇒ *bool* **where**
$irc \equiv irc\text{-}on\ UNIV$

lemma *irc-on-discard*:
assumes *irc-on A f*
assumes *finite C*
assumes $B \cup C \subseteq A$
assumes $f\ (B \cup C) \cap C = \{\}$
shows $f\ (B \cup C) = f\ B$

An equivalent condition is called *consistency* by some (Chambers and Yenmez (2013, Definition 2), Fleiner (2002, Equation (14))). Like *iia*, this formulation generalizes to infinite universes.

definition *consistency-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
$consistency\text{-}on\ A\ f \longleftrightarrow (\forall B{\subseteq}A.\ \forall C{\subseteq}B.\ f\ B \subseteq C \longrightarrow f\ B = f\ C)$

abbreviation *consistency* :: *'a cfun* ⇒ *bool* **where**
$consistency \equiv consistency\text{-}on\ UNIV$

lemma *irc-on-consistency-on*:
assumes *irc-on A f*
assumes *finite A*
shows *consistency-on A f*

lemma *consistency-on-irc-on*:
assumes *f-range-on A f*
assumes *consistency-on A f*
shows *irc-on A f*

These conditions imply that f is idempotent:

lemma *consistency-on-f-idem*:
assumes *f-range-on A f*
assumes *consistency-on A f*
assumes $B \subseteq A$
shows $f\ (f\ B) = f\ B$

Moulin (1985, p154) defines a similar but weaker property he calls *Aizerman*:

definition *Aizerman-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
$Aizerman\text{-}on\ A\ f \longleftrightarrow (\forall B{\subseteq}A.\ \forall C{\subseteq}B.\ f\ B \subseteq C \longrightarrow f\ C \subseteq f\ B)$

abbreviation *Aizerman* :: *'a cfun* ⇒ *bool* **where**
$Aizerman \equiv Aizerman\text{-}on\ UNIV$

lemma *consistency-on-Aizerman-on*:
assumes *consistency-on A f*
shows *Aizerman-on A f*

The converse requires f to be idempotent (Moulin 1985, p157):

lemma *Aizerman-on-idem-on-consistency-on*:
assumes *Aizerman-on A f*
assumes $\forall B{\subseteq}A.\ f\ (f\ B) = f\ B$
shows *consistency-on A f*

4.3 The *law of aggregate demand* condition aka *size monotonicity*

Hatfield and Milgrom (2005, §III) impose the *law of aggregate demand* (aka *size monotonicity*) to obtain the rural hospitals theorem (§5.6). It captures the following intuition:

> [...] Roughly, this law states that as the price falls, agents should demand more of a good. Here, price falls correspond to more contracts being available, and more demand corresponds to taking on (weakly) more contracts.

The *card* function takes a finite set into its cardinality (as a natural number).

definition *lad-on* :: *'a set* ⇒ *'a::finite cfun* ⇒ *bool* **where**
  lad-on A f ⟷ (∀ *B*⊆*A*. ∀ *C*⊆*B*. *card* (*f C*) ≤ *card* (*f B*))

abbreviation *lad* :: *'a::finite cfun* ⇒ *bool* **where**
  lad ≡ *lad-on UNIV*

This definition is identical amongst Hatfield and Milgrom (2005, §III), Fleiner (2002, (20)), and Aygün and Sönmez (2012b, Definition 4).

⟨*proof*⟩⟨*proof*⟩

Aygün and Sönmez (2012b, §5, Proposition 1) show that *substitutes* and *lad* imply *irc*, which therefore rescues many results in the matching-with-contracts literature.

lemma *lad-on-substitutes-on-irc-on*:
  assumes *f-range-on A f*
  assumes *substitutes-on A f*
  assumes *lad-on A f*
  shows *irc-on A f*

The converse does not hold.

4.4 The *expansion* condition

According to Moulin (1985, p152), a choice function satifies *expansion* if an alternative chosen from two sets is also chosen from their union.

definition *expansion-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
  expansion-on A f ⟷ (∀ *B*⊆*A*. ∀ *C*⊆*A*. *f B* ∩ *f C* ⊆ *f* (*B* ∪ *C*))

abbreviation *expansion* :: *'a cfun* ⇒ *bool* **where**
  expansion ≡ *expansion-on UNIV*

Condition γ due to Sen (1971) generalizes *expansion* to collections of sets of choices.

definition *expansion-gamma-on* :: *'a set* ⇒ *'a set set* ⇒ *'a cfun* ⇒ *bool* **where**
  expansion-gamma-on A As f ⟷ (⋃ *As*⊆*A* ∧ *As* ≠ {} ⟶ (⋂ *A*∈*As*. *f A*) ⊆ *f* (⋃ *As*))

definition *expansion-gamma* :: *'a set set* ⇒ *'a cfun* ⇒ *bool* **where**
  expansion-gamma ≡ *expansion-gamma-on UNIV*

lemma *expansion-gamma-expansion*:
  assumes ∀ *As*. *expansion-gamma-on A As f*
  shows *expansion-on A f*

lemma *expansion-expansion-gamma*:
  assumes *expansion-on A f*
  assumes *finite As*
  shows *expansion-gamma-on A As f*

The *expansion* condition plays a major role in the study of the *rationalizability* of choice functions, which we explore next.

4.5 Axioms of revealed preference

We digress from our taxonomy of conditions on choice functions to discuss *rationalizability*. A choice function is *rationalizable* if there exists some binary relation that generates it, typically by taking the *greatest* or *maximal* elements of the given set of alternatives:

definition *greatest* :: *'a rel* ⇒ *'a cfun* **where**
  greatest r X = $\{x{\in}X.\ \forall\, y{\in}X.\ (y,\ x) \in r\}$

definition *maximal* :: *'a rel* ⇒ *'a cfun* **where**
  maximal r X = $\{x{\in}X.\ \forall\, y{\in}X.\ \neg(x,\ y) \in r\}$

lemma (**in** *MaxR*) *greatest*:
  shows *set-option* (*MaxR-opt X*) = *greatest r* ($X \cap$ *Field r*)
⟨*proof*⟩⟨*proof*⟩

Note that *greatest* requires the relation to be reflexive and total, and *maximal* requires it to be irreflexive, for the choice functions to ever yield non-empty sets.

This game of uncovering the preference relations (if any) underlying a choice function goes by the name of *revealed preference*. (In contrast, later we show how these conditions guarantee the existence of stable many-to-one matches.) See Moulin (1985) and Border (2012) for background, intuition and critique, and Sen (1971) for further classical results and proofs.

We adopt the following notion here:

definition *rationalizes-on* :: *'a set* ⇒ *'a cfun* ⇒ *'a rel* ⇒ *bool* **where**
  rationalizes-on A f r ⟷ ($\forall\, B{\subseteq}A.\ f\ B$ = *greatest r B*)

abbreviation *rationalizes* :: *'a cfun* ⇒ *'a rel* ⇒ *bool* **where**
  rationalizes ≡ *rationalizes-on UNIV*

In words, relation r rationalizes the choice function f over universe A if $f\ B$ picks out the *greatest* elements of $B \subseteq A$ with respect to r. At this point r can be any relation that does the job, but soon enough we will ask that it satisfy some familiar ordering properties.

The analysis begins by determining under what constraints f can be rationalized, continues by establishing some properties of all rationalizable choice functions, and concludes by considering what it takes to establish stronger properties.

Following Border (2012, §5, Definition 2) and Sen (1971, Definition 2), we can generate the *revealed weakly preferred* relation for the choice function f:

definition *rwp-on* :: *'a set* ⇒ *'a cfun* ⇒ *'a rel* **where**
  rwp-on A f = $\{(x,\ y).\ \exists\, B{\subseteq}A.\ x \in B \wedge y \in f\ B\}$

abbreviation *rwp* :: *'a cfun* ⇒ *'a rel* **where**
  rwp ≡ *rwp-on UNIV*

lemma *rwp-on-refl-on*:
  assumes *f-range-on A f*
  assumes *decisive-on A f*
  shows *refl-on A* (*rwp-on A f*)

In words, if it is ever possible that $x \in B$ is available and $f\ B$ chooses y, then y is taken to always be at least as good as x.

The *V-axiom* asserts that whatever is revealed to be at least as good as anything else on offer is chosen:

definition *V-axiom-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**

$V\text{-}axiom\text{-}on\ A\ f \longleftrightarrow (\forall B \subseteq A.\ \forall y \in B.\ (\forall x \in B.\ (x,\ y) \in rwp\text{-}on\ A\ f) \longrightarrow y \in f\ B)$

abbreviation *V-axiom* :: *'a cfun* ⇒ *bool* **where**
V-axiom ≡ *V-axiom-on UNIV*

This axiom characterizes rationality; see Border (2012, Theorem 7). Sen (1971, §3) calls a decisive choice function that satisfies *V-axiom normal*.

lemma *rationalizes-on-f-range-on-V-axiom-on*:
assumes *rationalizes-on A f r*
shows *f-range-on A f*
and *V-axiom-on A f*

lemma *f-range-on-V-axiom-on-rationalizes-on*:
assumes *f-range-on A f*
assumes *V-axiom-on A f*
shows *rationalizes-on A f (rwp-on A f)*

theorem *V-axiom-on-rationalizes-on*:
shows $(f\text{-}range\text{-}on\ A\ f \wedge V\text{-}axiom\text{-}on\ A\ f) \longleftrightarrow (\exists r.\ rationalizes\text{-}on\ A\ f\ r)$

We could also ask that f be determined directly by how it behaves on pairs (Sen (1971), Moulin (1985, p151)), which turns out to be equivalent:

definition *rationalizable-binary-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
$rationalizable\text{-}binary\text{-}on\ A\ f \longleftrightarrow (\forall B \subseteq A.\ f\ B = \{y \in B.\ \forall x \in B.\ y \in f\ \{x,\ y\}\})$

abbreviation *rationalizable-binary* :: *'a cfun* ⇒ *bool* **where**
rationalizable-binary ≡ *rationalizable-binary-on UNIV*

theorem *V-axiom-realizable-binary*:
assumes *f-range-on A f*
shows *V-axiom-on A f* ⟷ *rationalizable-binary-on A f*⟨*proof*⟩

All rationalizable choice functions satisfy *iia* and *expansion* (Sen (1971), Moulin (1985, p152)).

lemma *rationalizable-binary-on-iia-on*:
assumes *f-range-on A f*
assumes *rationalizable-binary-on A f*
shows *iia-on A f*

lemma *rationalizable-binary-on-expansion-on*:
assumes *f-range-on A f*
assumes *rationalizable-binary-on A f*
shows *expansion-on A f*

The converse requires the set of alternatives to be finite, and moreover fails if the choice function is not *decisive*.

lemma *rationalizable-binary-on-converse*:
fixes *f* :: *'a::finite cfun*
assumes *f-range-on A f*
assumes *decisive-on A f*
assumes *iia-on A f*
assumes *expansion-on A f*
shows *rationalizable-binary-on A f*

That settles the issue of existence, but it is not clear that the relation is really "rational" (for instance, *rwp-on A f* need not be transitive). Therefore the analysis continues by further constraining the choice function so that it is rationalized by familiar ordering relations.

For instance, the following shows that the *axioms of revealed preference* are rationalized by total preorders (Sen 1971, Definitions 8 and 13)[1]. These are alo equivalent to some congruence axioms due to Samuelson (Border 2012). We define x to be *strictly revealed-preferred to* y if there is a situation where both are on offer and only y is chosen:

definition *rsp-on* :: *'a set* ⇒ *'a cfun* ⇒ *'a rel* **where** — (Sen 1971, Definition 8)
 $rsp\text{-}on\ A\ f = \{(x,\ y).\ \exists B{\subseteq}A.\ x \in Rf\ f\ B \wedge y \in f\ B\}$

abbreviation *rsp* :: *'a cfun* ⇒ *'a rel* **where**
 $rsp \equiv rsp\text{-}on\ UNIV$

This relation is typically denoted by P, for strict preference. The not-worse-than relation R is recovered by:

definition *rspR-on* :: *'a set* ⇒ *'a cfun* ⇒ *'a rel* **where** — (Sen 1971, Definition 9)
 $rspR\text{-}on\ A\ f = \{(x,\ y).\ \{x,\ y\} \subseteq A \wedge (y,\ x) \notin rsp\text{-}on\ A\ f\}$

abbreviation *rspR* :: *'a cfun* ⇒ *'a rel* **where**
 $rspR \equiv rspR\text{-}on\ UNIV$

Sen (1971, p309) defines the *weak axiom of revealed preference* (WARP) as follows:

definition *warp-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
 $warp\text{-}on\ A\ f \longleftrightarrow (\forall (x,\ y){\in}rsp\text{-}on\ A\ f.\ (y,\ x) \notin rwp\text{-}on\ A\ f)$

abbreviation *warp* :: *'a cfun* ⇒ *bool* **where**
 $warp \equiv warp\text{-}on\ UNIV$

The *strong axiom of revealed preference* (SARP) is essentially the transitive closure of *warp* (Sen 1971, p309):

definition *sarp-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
 $sarp\text{-}on\ A\ f \longleftrightarrow (\forall (x,\ y){\in}(rsp\text{-}on\ A\ f)^{+}.\ (y,\ x) \notin rwp\text{-}on\ A\ f)$

abbreviation *sarp* :: *'a cfun* ⇒ *bool* **where**
 $sarp \equiv sarp\text{-}on\ UNIV$

lemma *sarp-on-warp-on*: — Sen (1970, T.3 part)
 assumes *sarp-on A f*
 shows *warp-on A f*

lemma *rsp-on-irrefl*:
 $A \neq \{\} \Longrightarrow irrefl\ (rsp\text{-}on\ A\ f)$

For decisive choice functions, *warp* implies *sarp*. We show this following Sen (1971), via the *weak congruence axiom* (WCA): if f chooses x from some set B and y is revealed to be weakly preferred, then f must choose y from B as well.

definition *wca-on* :: *'a set* ⇒ *'a cfun* ⇒ *bool* **where**
 $wca\text{-}on\ A\ f \longleftrightarrow (\forall (x,\ y){\in}rwp\text{-}on\ A\ f.\ \forall B{\subseteq}A.\ x \in f\ B \wedge y \in B \longrightarrow y \in f\ B)$

abbreviation *wca* :: *'a cfun* ⇒ *bool* **where**
 $wca \equiv wca\text{-}on\ UNIV$

Decisive choice functions that satisfy *wca* are rationalized by total preorders, in particular *rwp*, and the converse obtains if they are normal.

lemma *wca-on-V-axiom-on*:
 assumes *wca-on A f*
 assumes *f-range-on A f*
 assumes *decisive-on A f*

[1]For Sen (1970, p9), an ordering is complete (total), reflexive, and transitive. Alternative names are: complete pre-ordering, complete quasi-ordering, and weak ordering.

shows *V-axiom-on A f*

lemma *wca-on-total-on*:
assumes *wca-on A f*
assumes *f-range-on A f*
assumes *decisive-on A f*
shows *total-on A (rwp-on A f)*

lemma *rwp-on-trans*:
assumes *wca-on A f*
assumes *f-range-on A f*
assumes *decisive-on A f*
shows *trans (rwp-on A f)*

lemma *wca-on-V-axiom-on-preorder-on*: — Sen (1970, T.1, T.3 part)
assumes *f-range-on A f*
assumes *decisive-on A f*
shows $wca\text{-}on\ A\ f \longleftrightarrow V\text{-}axiom\text{-}on\ A\ f \land preorder\text{-}on\ A\ (rwp\text{-}on\ A\ f) \land total\text{-}on\ A\ (rwp\text{-}on\ A\ f)\langle proof\rangle$

lemma *wca-on-rwp-on-rspR-on*: — Sen (1970, T.2)
assumes *wca-on A f*
assumes *f-range-on A f*
assumes *decisive-on A f*
shows $rwp\text{-}on\ A\ f = rspR\text{-}on\ A\ f\langle proof\rangle$

lemma *rwp-on-rspR-on-wca-on*: — Sen (1970, T.2)
assumes $rwp\text{-}on\ A\ f = rspR\text{-}on\ A\ f$
shows *wca-on A f*

lemma *wca-on-warp-on*: — Sen (1970, T.3 part)
shows $wca\text{-}on\ A\ f \longleftrightarrow warp\text{-}on\ A\ f$

lemma *warp-on-sarp-on*: — Sen (1970, T.3 part)
assumes *warp-on A f*
assumes *f-range-on A f*
assumes *decisive-on A f*
shows *sarp-on A f*
$\langle proof\rangle$
$\langle proof\rangle$

The *decisive* constraint here is necessary: consider a Condorcet cycle over $\{x, y, z\}$: forcing $f\ \{x, y, z\}$ to be non-empty resolves this.

Sen (1971) proves that these and other conditions on choice functions are equivalent (under the *decisive* hypothesis).

4.5.1 The *strong axiom of revealed preference* ala Aygün and Sönmez (2012b)

Aygün and Sönmez (2012b, §6) adopt a different definition for a *strong axiom of revealed preference* and show that it holds for all choice functions that satisfy *iia* and *consistency*.

abbreviation *nth-mod* :: *'a list* ⇒ *nat* ⇒ *'a* (**infixl** ‹!%› *100*) **where**
$xs\ !\%\ i \equiv xs\ !\ (i\ mod\ length\ xs)$

definition *mwc-sarp* :: *'a cfun* ⇒ *bool* **where**
$mwc\text{-}sarp\ f \longleftrightarrow$
$\neg(\exists Xs.\ length\ Xs > 1 \land distinct\ (map\ f\ Xs) \land (\forall i.\ f\ (Xs!\%i) \subset Xs!\%i \cap Xs!\%(i{+}1)))$

lemma *iia-consistency-mwc-sarp*:
assumes *f-range f*
assumes *iia f* — *substitutes*

assumes *consistency f — irc*
shows *mwc-sarp f*
⟨*proof*⟩
⟨*proof*⟩

4.6 Choice functions arising from linear orders

An obvious way to construct a choice function is to derive one from a linear order, i.e., a list of strict preferences. We allow such rankings to omit some alternatives, which means the resulting function is not decisive.

We work with a finite universe here.

locale *linear-cf* =
fixes *r* :: *'a::finite rel*
fixes *linear-cf* :: *'a cfun*
assumes *r-linear*: *Linear-order r*
assumes *linear-cf-def*: *linear-cf X* ≡ *set-option* (*MaxR.MaxR-opt r X*)
begin

interpretation *MaxR*: *MaxR r* ⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
lemma *range*:
shows *linear-cf X* ⊆ *X* ∩ *Field r*

lemmas *range'* = *rev-subsetD*[*OF - range*, *of x*] **for** *x*

lemma *singleton*:
shows $x \in$ *linear-cf X* ⟷ *linear-cf X* = $\{x\}$

lemma *subset*:
assumes *linear-cf Y* ⊆ *X*
assumes *X* ⊆ *Y*
shows *linear-cf Y* = *linear-cf X*

lemma *union*:
shows *linear-cf* (*X* ∪ *Y*) = (*if linear-cf X* = {} *then linear-cf Y else if linear-cf Y* = {} *then linear-cf X else* {*MaxR.maxR x y* |*x y*. *x* ∈ *linear-cf X* ∧ *y* ∈ *linear-cf Y*})

lemma *mono*:
assumes *x* ∈ *linear-cf X*
shows ∃ *y* ∈ *linear-cf* (*X* ∪ *Y*). (*x*, *y*) ∈ *r*

lemmas *greatest* = *MaxR.greatest*[*folded linear-cf-def*]

lemma *preferred*:
assumes (*x*, *y*) ∈ *r*
assumes *x* ∈ *linear-cf X*
assumes *y* ∈ *X*
shows *y* = *x*

lemma *card-le*:
shows *card* (*linear-cf X*) ≤ *1*

lemma *card*:
shows *card* (*linear-cf X*) = (*if X* ∩ *Field r* = {} *then 0 else 1*)

lemma *f-range*:
shows *f-range-on X linear-cf*

lemma *domain*:

shows *linear-cf* $(X \cap \mathit{Field}\ r) = \mathit{linear\text{-}cf}\ X$

lemma *decisive-on*:
shows *decisive-on* $(\mathit{Field}\ r)$ *linear-cf*

lemma *resolute-on*:
shows *resolute-on* $(\mathit{Field}\ r)$ *linear-cf*

lemma *Rf-mono-on*:
shows *mono-on* X $(\mathit{Rf}\ \mathit{linear\text{-}cf})$

lemmas $\mathit{iia} = \mathit{iffD1}[\mathit{OF}\ \mathit{Rf\text{-}mono\text{-}on\text{-}iia\text{-}on}\ \mathit{Rf\text{-}mono\text{-}on}]$

lemma *Chernoff*:
shows *Chernoff-on* X *linear-cf*

lemma *irc*:
shows *irc-on* X *linear-cf*

lemma *consistency*:
shows *consistency-on* X *linear-cf*

lemma *lad*:
shows *lad-on* X *linear-cf*

end

4.7 Plott's *path independence* condition

As recognised by Fleiner (2002, §4) and Chambers and Yenmez (2013) in the context of matching with contracts, the *irc* and *substitutes* conditions together are equivalent to *path independence*, a condition introduced to the social choice setting by Plott (1973). Moulin (1985, Lemma 6) ascribes this equivalence result to Aizerman and Malishevski (1981).

definition *path-independent-on* :: $'a\ \mathit{set} \Rightarrow {'a}\ \mathit{cfun} \Rightarrow \mathit{bool}$ **where**
$\mathit{path\text{-}independent\text{-}on}\ A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) = f\ (B \cup f\ C))$

abbreviation *path-independent* :: $'a\ \mathit{cfun} \Rightarrow \mathit{bool}$ **where**
$\mathit{path\text{-}independent} \equiv \mathit{path\text{-}independent\text{-}on}\ \mathit{UNIV}$

Intuitively a choice function satisfying this condition ignores the order in which choices are made in the following sense:

lemma *path-independent-on-symmetric*:
assumes *f-range-on* $A\ f$
shows $\mathit{path\text{-}independent\text{-}on}\ A\ f \longleftrightarrow (\forall B\ C.\ B \subseteq A \wedge C \subseteq A \longrightarrow f\ (B \cup C) = f\ (f\ B \cup f\ C))$

lemma *path-independent-on-Chernoff-on*:
assumes *path-independent-on* $A\ f$
assumes *f-range-on* $A\ f$
shows *Chernoff-on* $A\ f$

lemma *path-independent-on-consistency-on*:
assumes *path-independent-on* $A\ f$
shows *consistency-on* $A\ f$

lemma *Chernoff-on-consistency-on-path-independent-on*:
assumes *f-range-on* $A\ f$
shows $\mathit{Chernoff\text{-}on}\ A\ f \wedge \mathit{consistency\text{-}on}\ A\ f \longleftrightarrow \mathit{path\text{-}independent\text{-}on}\ A\ f$

lemma (**in** *linear-cf*) *path-independent*:
  shows *path-independent linear-cf*

4.7.1 Path independence and decomposition into orderings

We now show that a choice function over a finite universe satisfying *path-independent* is characterized by taking the maximum elements of some finite set of orderings.

Moulin (1985, Definition 12) says that a choice function is *pseudo-rationalized* by the orderings *Rs* if *f* chooses all of the *greatest* *r* elements of *B* for each $r \in Rs$:

definition *pseudo-rationalizable-on* :: $'a$::*finite set* $\Rightarrow$ $'a$ *rel set* $\Rightarrow$ $'a$ *cfun* $\Rightarrow$ *bool* **where**
  pseudo-rationalizable-on A Rs f
    $\longleftrightarrow (\forall\, r{\in}Rs.\ Linear\text{-}order\ r) \wedge (\forall\, B{\subseteq}A.\ f\ B = (\bigcup r{\in}Rs.\ greatest\ r\ (B \cap Field\ r)))$

lemma *pseudo-rationalizable-on-def2*:
  pseudo-rationalizable-on A Rs f
    $\longleftrightarrow (\forall\, r{\in}Rs.\ Linear\text{-}order\ r) \wedge (\forall\, B{\subseteq}A.\ f\ B = (\bigcup r{\in}Rs.\ set\text{-}option\ (MaxR.MaxR\text{-}opt\ r\ B)))$

We deviate from Moulin in using non-total linear orders, where his are total, asymmetric, and transitive; in other words, strict total linear orders. This allows us to treat non-decisive choice functions, and we later show that the choice function is decisive iff the orders are total.

Moulin (1985, Theorem 5) assumes *Aizerman* and *Chernoff*, which are equivalent to *path-independent*.

lemma *Aizerman-on-Chernoff-on-path-independent-on*:
  assumes *f-range-on A f*
  shows *Aizerman-on A f* $\wedge$ *Chernoff-on A f* $\longleftrightarrow$ *path-independent-on A f*

It is straightforward to show that pseudo-rationalizable choice functions satisfy *path-independent* using the properties of *MaxR.MaxR-opt*:

lemma *pseudo-rationalizable-on-path-independent-on*:
  assumes *pseudo-rationalizable-on A Rs f*
  shows *path-independent-on A f*

The converse requires that we construct a suitable set of orderings that rationalize $f\ C$ for each $C \subseteq A$. We do this by finding a set $B \subseteq A$ where $f\ B \subseteq C$ by successively removing elements in $f\ A - f\ C$. (As these elements are chosen by f from supersets of B, we rank these above all of those in $f\ B$.) By *consistency* (§4.2), $f\ C = f\ B$. We generate one order for each element of $f\ C$. Some extra care takes care of *decisive* choice functions.

Termination is guaranteed by the finiteness of A and the *f-range-on* hypothesis.

context
  fixes A :: $'a$::*finite set*
  fixes f :: $'a$ *cfun*
  notes *conj-cong*[*fundef-cong*]
begin

function (*domintros*) *mk-linear-orders* :: $'a$ *set* $\Rightarrow$ $'a$ *set* $\Rightarrow$ $'a$ *list set* **where**
  mk-linear-orders C B =
   (*if* $f\ B = \{\}$ *then* $\{[]\}$
    else if $f\ B \subseteq C$
        then $\{b \# cs\ |b\ cs.\ b \in f\ B \wedge cs \in mk\text{-}linear\text{-}orders\ \{\}\ (B - \{b\})\}$
        else let $b = SOME\ x.\ x \in f\ B - C$ *in* $\{b \# cs\ |cs.\ cs \in mk\text{-}linear\text{-}orders\ C\ (B - \{b\})\})$

context
  assumes *f-range-on A f*
begin
⟨*proof*⟩⟨*proof*⟩

lemma *mk-linear-orders-non-empty*:
assumes $B \subseteq A$
shows $\exists\, r.\ r \in$ *mk-linear-orders* $C\ B$

lemma *mk-linear-orders-range*:
assumes $r \in$ *mk-linear-orders* $C\ B$
assumes $B \subseteq A$
shows *set* $r \subseteq B$

lemma *mk-linear-orders-nth*:
assumes $r \in$ *mk-linear-orders* $C\ B$
assumes $B \subseteq A$
assumes $i <$ *length* r
shows $r\ !\ i \in f\ (B - \mathit{set}\ (\mathit{take}\ i\ r))$

lemma *mk-linear-orders-distinct*:
assumes $r \in$ *mk-linear-orders* $C\ B$
assumes $B \subseteq A$
shows *distinct* r

lemma *mk-linear-orders-Linear-order*:
assumes $r \in$ *mk-linear-orders* $C\ A$
shows *Linear-order* (*linord-of-list* r)

lemma *mk-linear-orders-decisive-on-set-r*:
assumes $r \in$ *mk-linear-orders* $C\ B$
assumes *decisive-on* $A\ f$
assumes $B \subseteq A$
shows *set* $r = B$

lemma *mk-linear-orders-decisive-on-refl-on*:
assumes $r \in$ *mk-linear-orders* $C\ A$
assumes *decisive-on* $A\ f$
shows *refl-on* A (*linord-of-list* r)

lemma *mk-linear-orders-decisive-on-total-on*:
assumes $r \in$ *mk-linear-orders* $C\ A$
assumes *decisive-on* $A\ f$
shows *total-on* A (*linord-of-list* r)

lemma *mk-linear-orders-set-r-decisive-on*:
assumes $r \in$ *mk-linear-orders* $C\ B$
assumes $B \subseteq A$
assumes $B \subseteq$ *set* r
assumes *iia-on* $A\ f$
shows *decisive-on* $B\ f$

lemma *mk-linear-orders-total-on-decisive-on*:
assumes $r \in$ *mk-linear-orders* $C\ A$
assumes $A \subseteq$ *set* r
assumes *iia-on* $A\ f$
shows *decisive-on* $A\ f$

lemma *mk-linear-orders-MaxR-opt-f*:
assumes $r \in$ *mk-linear-orders* $C\ A$
assumes *MaxR.MaxR-opt* (*linord-of-list* r) $D =$ *Some* x
assumes *iia-on* $A\ f$
assumes $D \subseteq A$

shows $x \in f\ D$

lemma *mk-linear-orders-f-MaxR-opt*:
assumes $x \in f\ C$
assumes *consistency-on* $A\ f$
assumes $B \subseteq A$
assumes $C \subseteq B$
shows $\exists\, r{\in}\mathit{mk\text{-}linear\text{-}orders}\ C\ B.\ \mathit{MaxR.MaxR\text{-}opt}\ (\mathit{linord\text{-}of\text{-}list}\ r)\ C = \mathit{Some}\ x$

end

end

lemma *path-independent-on-pseudo-rationalizable-on*:
fixes $f :: {}'a{::}\mathit{finite}\ \mathit{cfun}$
assumes *path-independent-on* $A\ f$
assumes *f-range-on* $A\ f$
assumes *Rs-def*[*simp*]: $\mathit{Rs} = (\bigcup C{\in}\mathit{Pow}\ A.\ \mathit{linord\text{-}of\text{-}list}\ \text{`}\ \mathit{mk\text{-}linear\text{-}orders}\ f\ C\ A)$
shows $\mathit{pseudo\text{-}rationalizable\text{-}on}\ A\ \mathit{Rs}\ f \wedge (\forall\, r{\in}\mathit{Rs}.\ \mathit{refl\text{-}on}\ A\ r \wedge \mathit{total\text{-}on}\ A\ r \longleftrightarrow \mathit{decisive\text{-}on}\ A\ f)$

Our top-level theorem is essentially Moulin (1985, Theorem 5):

theorem *pseudo-rationalizable*:
assumes *f-range-on* $A\ f$
shows *path-independent-on* $A\ f$
$\longleftrightarrow (\exists\, \mathit{Rs}.\ \mathit{pseudo\text{-}rationalizable\text{-}on}\ A\ \mathit{Rs}\ f \wedge (\forall\, r{\in}\mathit{Rs}.\ \mathit{refl\text{-}on}\ A\ r \wedge \mathit{total\text{-}on}\ A\ r \longleftrightarrow \mathit{decisive\text{-}on}\ A\ f))$

5 Hatfield and Milgrom (2005): Matching with contracts

We take the original paper on matching with contracts by Hatfield and Milgrom (2005) as our roadmap, which follows a similar path to Roth and Sotomayor (1990, §2.5). We defer further motivation to these texts. Our first move is to capture the scenarios described in their §I(A) (p916) in a locale.

locale *Contracts* =
fixes $\mathit{Xd} :: {}'x{::}\mathit{finite} \Rightarrow {}'d{::}\mathit{finite}$
fixes $\mathit{Xh} :: {}'x \Rightarrow {}'h{::}\mathit{finite}$
fixes $\mathit{Pd} :: {}'d \Rightarrow {}'x\ \mathit{rel}$
fixes $\mathit{Ch} :: {}'h \Rightarrow {}'x\ \mathit{cfun}$
assumes *Pd-linear*: $\forall\, d.\ \mathit{Linear\text{-}order}\ (\mathit{Pd}\ d)$
assumes *Pd-range*: $\forall\, d.\ \mathit{Field}\ (\mathit{Pd}\ d) \subseteq \{x.\ \mathit{Xd}\ x = d\}$
assumes *Ch-range*: $\forall\, h.\ \forall\, X.\ \mathit{Ch}\ h\ X \subseteq \{x{\in}X.\ \mathit{Xh}\ x = h\}$
assumes *Ch-singular*: $\forall\, h.\ \forall\, X.\ \mathit{inj\text{-}on}\ \mathit{Xd}\ (\mathit{Ch}\ h\ X)$
begin

The set of contracts is modelled by the type $'x$, a free type variable that will later be interpreted by some non-empty set. Similarly $'d$ and $'h$ track the names of doctors and hospitals respectively. All of these are finite by virtue of belonging to the *finite* type class.
We fix four constants:

- *Xd* (*Xh*) projects the name of the relevant doctor (hospital) from a contract;
- *Pd* maps doctors to their linear preferences over some subset of contracts that name them (assumptions *Pd-linear* and *Pd-range*); and
- *Ch* maps hospitals to their choice functions (§4), which are similarly constrained to contracts that name them (assumption *Ch-range*). Moreover their choices must name each doctor at most once, i.e., *Xd* must be injective on these (assumption *Ch-singular*).

The reader familiar with the literature will note that we do not have a null contract (also said to represent the *outside option* of unemployment), and instead use partiality of the doctors' preferences. This provides two benefits: firstly, Xh is a total function here, and secondly we achieve some economy of description when instantiating this locale as Pd only has to rank the relevant contracts.

We note in passing that neither the doctors' nor hospitals' choice functions are required to be decisive, unlike the standard literature on choice functions (§4).

In addition to the above, the following constitute the definitions that must be trusted for the results we prove to be meaningful.

definition $Cd :: {}'d \Rightarrow {}'x\ cfun$ **where**
  $Cd\ d \equiv set\text{-}option \circ MaxR.MaxR\text{-}opt\ (Pd\ d)$

definition $CD\text{-}on :: {}'d\ set \Rightarrow {}'x\ cfun$ **where**
  $CD\text{-}on\ ds\ X = (\bigcup d{\in}ds.\ Cd\ d\ X)$

abbreviation $CD :: {}'x\ set \Rightarrow {}'x\ set$ **where**
  $CD \equiv CD\text{-}on\ UNIV$

definition $CH :: {}'x\ cfun$ **where**
  $CH\ X = (\bigcup h.\ Ch\ h\ X)$

The function Cd constructs a choice function from the doctor's linear preferences (see §4.6). Both CD and CH simply aggregate opinions in the obvious way. The functions $CD\text{-}on$ is parameterized with a set of doctors to support the proofs of §5.5.

We also define RD (Rh, RH) to be the subsets of a given set of contracts that are rejected by the doctors (hospitals). (The abbreviation Rf is defined in §4.)

abbreviation $(input)$ $RD\text{-}on :: {}'d\ set \Rightarrow {}'x\ cfun$ **where**
  $RD\text{-}on\ ds \equiv Rf\ (CD\text{-}on\ ds)$

abbreviation $(input)$ $RD :: {}'x\ cfun$ **where**
  $RD \equiv RD\text{-}on\ UNIV$

abbreviation $(input)$ $Rh :: {}'h \Rightarrow {}'x\ cfun$ **where**
  $Rh\ h \equiv Rf\ (Ch\ h)$

abbreviation $(input)$ $RH :: {}'x\ cfun$ **where**
  $RH \equiv Rf\ CH$

A *mechanism* maps doctor and hospital preferences into a match (here a set of contracts).

type-synonym (**in** $-$) $({}'d, {}'h, {}'x)\ mechanism = ({}'d \Rightarrow {}'x\ rel) \Rightarrow ({}'h \Rightarrow {}'x\ cfun) \Rightarrow {}'d\ set \Rightarrow {}'x\ set\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle pro$

An *allocation* is a set of contracts where each names a distinct doctor. (Hospitals can contract multiple doctors.)

abbreviation $(input)$ $allocation :: {}'x\ set \Rightarrow bool$ **where**
  $allocation \equiv inj\text{-}on\ Xd$

We often wish to extract a doctor's or a hospital's contract from an *allocation*.

definition $dX :: {}'x\ set \Rightarrow {}'d \Rightarrow {}'x\ set$ **where**
  $dX\ X\ d = \{x \in X.\ Xd\ x = d\}$

definition $hX :: {}'x\ set \Rightarrow {}'h \Rightarrow {}'x\ set$ **where**
  $hX\ X\ h = \{x \in X.\ Xh\ x = h\}\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle proof\rangle\langle proof\rangle$

Stability is the key property we look for in a match (here a set of contracts), and consists of two parts.

Firstly, we ask that it be *individually rational*, i.e., the contracts in the match are actually acceptable to all participants. Note that this implies the match is an *allocation*.

definition $individually\text{-}rational\text{-}on :: {}'d\ set \Rightarrow {}'x\ set \Rightarrow bool$ **where**
  $individually\text{-}rational\text{-}on\ ds\ X \longleftrightarrow CD\text{-}on\ ds\ X = X \wedge CH\ X = X$

abbreviation *individually-rational* :: *'x set* ⇒ *bool* **where**
individually-rational ≡ *individually-rational-on UNIV*

The second condition requires that there be no coalition of a hospital and one or more doctors who prefer another set of contracts involving them; the hospital strictly, the doctors weakly. Contrast this definition with the classical one for stable marriages given in §2.

definition *blocking-on* :: *'d set* ⇒ *'x set* ⇒ *'h* ⇒ *'x set* ⇒ *bool* **where**
$blocking\text{-}on\ ds\ X\ h\ X' \longleftrightarrow X' \neq Ch\ h\ X \wedge X' = Ch\ h\ (X \cup X') \wedge X' \subseteq CD\text{-}on\ ds\ (X \cup X')$

definition *stable-no-blocking-on* :: *'d set* ⇒ *'x set* ⇒ *bool* **where**
$stable\text{-}no\text{-}blocking\text{-}on\ ds\ X \longleftrightarrow (\forall\, h\ X'.\ \neg blocking\text{-}on\ ds\ X\ h\ X')$

abbreviation *stable-no-blocking* :: *'x set* ⇒ *bool* **where**
stable-no-blocking ≡ *stable-no-blocking-on UNIV*

definition *stable-on* :: *'d set* ⇒ *'x set* ⇒ *bool* **where**
$stable\text{-}on\ ds\ X \longleftrightarrow individually\text{-}rational\text{-}on\ ds\ X \wedge stable\text{-}no\text{-}blocking\text{-}on\ ds\ X$

abbreviation *stable* :: *'x set* ⇒ *bool* **where**
stable ≡ *stable-on UNIV*⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

end

5.1 Theorem 1: Existence of stable pairs

We proceed to define a function whose fixed points capture all stable matches. Hatfield and Milgrom (2005, I(B), p917) provide the following intuition:

> The first theorem states that a set of contracts is stable if any alternative contract would be rejected by some doctor or some hospital from its suitably defined opportunity set. In the formulas below, think of the doctors' opportunity set as *XD* and the hospitals' opportunity set as *XH*. If X' is the corresponding stable set, then *XD* must include, in addition to X', all contracts that would not be rejected by the hospitals, and *XH* must similarly include X' and all contracts that would not be rejected by the doctors. If X' is stable, then every alternative contract is rejected by somebody, so $X = XH \cup XD$ [where X is the set of all contracts]. This logic is summarized in the first theorem.

See also Fleiner (2003, p6,§4) and Fleiner (2002, §2), from whom we adopt the term *stable pair*.

context *Contracts*
begin

definition *stable-pair-on* :: *'d set* ⇒ *'x set* × *'x set* ⇒ *bool* **where**
$stable\text{-}pair\text{-}on\ ds = (\lambda(XD,\ XH).\ XD = -\ RH\ XH \wedge XH = -\ RD\text{-}on\ ds\ XD)$

abbreviation *stable-pair* :: *'x set* × *'x set* ⇒ *bool* **where**
stable-pair ≡ *stable-pair-on UNIV*

abbreviation *match* :: *'x set* × *'x set* ⇒ *'x set* **where**
$match\ X \equiv fst\ X \cap snd\ X$

Hatfield and Milgrom (2005, Theorem 1) state that every solution (*XD*, *XH*) of *stable-pair* yields a stable match $XD \cap XH$, and conversely, i.e., every stable match is the intersection of some stable pair. Aygün and Sönmez (2012b) show that neither is the case without further restrictions on the hospitals' choice functions *Ch*; we exhibit their counterexample below.
Even so we can establish some properties in the present setting:

lemma *stable-pair-on-CD-on*:
assumes *stable-pair-on ds XD-XH*
shows $match\ XD\text{-}XH = CD\text{-}on\ ds\ (fst\ XD\text{-}XH)$

lemma *stable-pair-on-CH*:
  assumes *stable-pair-on ds XD-XH*
  shows *match XD-XH* = *CH* (*snd XD-XH*)

lemma *stable-pair-on-CD-on-CH*:
  assumes *stable-pair-on ds XD-XH*
  shows *CD-on ds* (*fst XD-XH*) = *CH* (*snd XD-XH*)

lemma *stable-pair-on-allocation*:
  assumes *stable-pair-on ds XD-XH*
  shows *allocation* (*match XD-XH*)
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

We run out of steam on the following two lemmas, which are the remaining requirements for stability.

lemma
  assumes *stable-pair-on ds XD-XH*
  shows *individually-rational-on ds* (*match XD-XH*)
⟨*proof*⟩

lemma
  assumes *stable-pair-on ds XD-XH*
  shows *stable-no-blocking* (*match XD-XH*)
⟨*proof*⟩

Hatfield and Milgrom (2005) also claim that the converse holds:

lemma
  assumes *stable-on ds X*
  obtains *XD-XH* **where** *stable-pair-on ds XD-XH* **and** *X* = *match XD-XH*
⟨*proof*⟩

Again, the following counterexample shows that the *substitutes* condition on *Ch* is too weak to guarantee this. We show it holds under stronger assumptions in §5.1.3.

end

5.1.1 Theorem 1 does not hold (Aygün and Sönmez 2012b)

The following counterexample, due to Aygün and Sönmez (2012b, §3: Example 2), comprehensively demonstrates that Hatfield and Milgrom (2005, Theorem 1) does not hold.

We create three types: *D2* consists of two elements, representing the doctors, and *H* is the type of the single hospital. There are four contracts in the type *X4*.

datatype *D2* = *D1* | *D2*
datatype *H1* = *H*
datatype *X4* = *Xd1* | *Xd1'* | *Xd2* | *Xd2'*⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
primrec *X4d* :: *X4* ⇒ *D2* **where**
  X4d Xd1 = *D1*
| *X4d Xd1'* = *D1*
| *X4d Xd2* = *D2*
| *X4d Xd2'* = *D2*

abbreviation *X4h* :: *X4* ⇒ *H1* **where**
  X4h - ≡ *H*

primrec *PX4d* :: *D2* ⇒ *X4 rel* **where**
  PX4d D1 = *linord-of-list* [*Xd1'*, *Xd1*]
| *PX4d D2* = *linord-of-list* [*Xd2*, *Xd2'*]

function *CX4h* :: *H1* ⇒ *X4 cfun* **where**

$CX4h$ - $\{Xd1\} = \{Xd1\}$
| $CX4h$ - $\{Xd1'\} = \{Xd1'\}$
| $CX4h$ - $\{Xd2\} = \{Xd2\}$
| $CX4h$ - $\{Xd2'\} = \{Xd2'\}$
| $CX4h$ - $\{Xd1, Xd1'\} = \{Xd1\}$
| $CX4h$ - $\{Xd1, Xd2\} = \{Xd1, Xd2\}$
| $CX4h$ - $\{Xd1, Xd2'\} = \{Xd2'\}$
| $CX4h$ - $\{Xd1', Xd2\} = \{Xd1'\}$
| $CX4h$ - $\{Xd1', Xd2'\} = \{Xd1', Xd2'\}$
| $CX4h$ - $\{Xd2, Xd2'\} = \{Xd2\}$
| $CX4h$ - $\{Xd1, Xd1', Xd2\} = \{\}$
| $CX4h$ - $\{Xd1, Xd1', Xd2'\} = \{\}$
| $CX4h$ - $\{Xd1, Xd2, Xd2'\} = \{\}$
| $CX4h$ - $\{Xd1', Xd2, Xd2'\} = \{\}$
| $CX4h$ - $\{Xd1, Xd1', Xd2, Xd2'\} = \{\}$
| $CX4h$ - $\{\} = \{\}$
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

interpretation *StableNoDecomp*: *Contracts X4d X4h PX4d CX4h*

There are two stable matches in this model.

⟨*proof*⟩⟨*proof*⟩

lemma *stable*:
shows *StableNoDecomp.stable* $X \longleftrightarrow X = \{Xd1, Xd2\} \vee X = \{Xd1', Xd2'\}$⟨*proof*⟩

However neither arises from a pair XD, XH that satisfy *StableNoDecomp.stable-pair*:

lemma *StableNoDecomp-XD-XH*:
shows *StableNoDecomp.stable-pair* $(XD, XH) \longleftrightarrow (XD = \{\} \wedge XH = \{Xd1, Xd1', Xd2, Xd2'\})$⟨*proof*⟩

proposition
assumes *StableNoDecomp.stable-pair* (XD, XH)
shows $\neg$*StableNoDecomp.stable* $(XD \cap XH)$

Moreover the converse of Theorem 1 does not hold either: the single decomposition that satisfies *StableNoDecomp.stable-pair* (*StableNoDecomp-XD-XH*) does not yield a stable match:

proposition
assumes *StableNoDecomp.stable* X
shows $\neg(\exists XD\ XH.$ *StableNoDecomp.stable-pair* $(XD, XH) \wedge X = XD \cap XH)$

So there is not hope for Hatfield and Milgrom (2005, Theorem 1) as it stands. Note that the counterexample satisfies the *substitutes* condition (see §4.1):

lemma
shows *substitutes* $(CX4h\ H)$

Therefore while *substitutes* supports the monotonicity argument that underpins their deferred-acceptance algorithm (see §5.2), it is not enough to rescue Theorem 1. One way forward is to constrain the hospitals' choice functions, which we discuss in the next section.

5.1.2 Theorem 1 holds with *independence of rejected contracts*

Aygün and Sönmez (2012b) propose to rectify this issue by requiring hospitals' choices to satisfy *irc* (§4.2). Reassuringly their counterexample fails to satisfy it:

lemma

shows $\neg irc\ (CX4h\ H)$

We adopt this hypothesis by extending the *Contracts* locale:

locale *ContractsWithIRC* = *Contracts* +
assumes *Ch-irc*: $\forall h.\ irc\ (Ch\ h)$
begin

This property requires that *Ch* behave, for example, as follows:

lemma *Ch-domain*:
shows $Ch\ h\ (A \cap \{x.\ Xh\ x = h\}) = Ch\ h\ A$

lemmas *Ch-irc-idem* = *consistency-on-f-idem*[*OF Ch-f-range Ch-consistency*, *simplified*]

lemma *CH-irc-idem*:
shows $CH\ (CH\ A) = CH\ A$

lemma *Ch-CH-irc-idem*:
shows $Ch\ h\ (CH\ A) = Ch\ h\ A$

This suffices to show the left-to-right direction of Theorem 1.

lemma *stable-pair-on-individually-rational*:
assumes *stable-pair-on ds XD-XH*
shows *individually-rational-on ds* (*match XD-XH*)

lemma *stable-pair-on-stable-no-blocking-on*:
assumes *stable-pair-on ds XD-XH*
shows *stable-no-blocking-on ds* (*match XD-XH*)
⟨*proof*⟩

theorem *stable-pair-on-stable-on*:
assumes *stable-pair-on ds XD-XH*
shows *stable-on ds* (*match XD-XH*)

end

5.1.3 The converse of Theorem 1

The forward direction of Theorem 1 gives us a way of finding stable matches by computing fixed points of a function closely related to *stable-pair* (see §5.2). The converse says that every stable match can be decomposed in this way, which implies that the stable matches form a lattice (see also §5.2).

The following proofs assume that the hospitals' choice functions satisfy *substitutes* and *irc*.

context *ContractsWithIRC*
begin

context
fixes *ds* :: *'b set*
fixes *X* :: *'a set*
begin

Following Hatfield and Milgrom (2005, Proof of Theorem 1), we partition the set of all contracts into [X, *XD-smallest* − X, *XH-largest* − X] with careful definitions of the two sets *XD-smallest* and *XH-largest*. Specifically *XH-largest* contains all contracts ranked at least as good as those in X by the doctors, considering unemployment and unacceptable contracts. Similarly *XD-smallest* contains those ranked at least as poorly.

definition *XH-largest* :: *'a set* **where**
XH-largest =

$\{y.\ Xd\ y \in ds$
$\wedge\ y \in Field\ (Pd\ (Xd\ y))$
$\wedge\ (\forall\, x \in dX\ X\ (Xd\ y).\ (x,\ y) \in Pd\ (Xd\ y))\}$

definition *XD-smallest* :: *'a set* **where**
XD-smallest = − (*XH-largest* − *X*)

context
assumes *stable-on ds X*
begin

lemma *Ch-XH-largest-Field*:
assumes $x \in$ *Ch h XH-largest*
shows $x \in$ *Field* (*Pd* (*Xd x*))
⟨*proof*⟩

lemma *Ch-XH-largest-Xd*:
assumes $x \in$ *Ch h XH-largest*
shows *Xd x* $\in$ *ds*
⟨*proof*⟩

lemma *X-subseteq-XH-largest*:
shows $X \subseteq$ *XH-largest*
⟨*proof*⟩

lemma *X-subseteq-XD-smallest*:
shows $X \subseteq$ *XD-smallest*
⟨*proof*⟩

lemma *X-XD-smallest-XH-largest*:
shows $X =$ *XD-smallest* $\cap$ *XH-largest*
⟨*proof*⟩

The goal of the next few lemmas is to show the constituents of *stable-pair-on ds* (*XD-smallest*, *XH-largest*). Intuitively, if a doctor has a contract x in X, then all of their contracts in *XH-largest* are at least as desirable as x, and so the *stable-no-blocking* hypothesis guarantees the hospitals choose x from *XH-largest*, and similarly the doctors x from *XD-smallest*.

lemma *XH-largestCdXXH-largest*:
assumes $x \in$ *Ch h XH-largest*
shows $x \in$ *Cd* (*Xd x*) ($X \cup$ *Ch h XH-largest*)
⟨*proof*⟩

lemma *CH-XH-largest*:
shows *CH XH-largest* = *X*
⟨*proof*⟩

lemma *Cd-XD-smallest*:
assumes $d \in ds$
shows *Cd d* (*XD-smallest* $\cap$ *Field* (*Pd d*)) = *Cd d* ($X \cap$ *Field* (*Pd d*))
⟨*proof*⟩

lemma *CD-on-XD-smallest*:
shows *CD-on ds XD-smallest* = *X*
⟨*proof*⟩

theorem *stable-on-stable-pair-on*:
shows *stable-pair-on ds* (*XD-smallest*, *XH-largest*)
⟨*proof*⟩

end

end

Our ultimate statement of Theorem 1 of Hatfield and Milgrom (2005) ala Aygün and Sönmez (2012b) goes as follows, bearing in mind that we are working in the *ContractsWithIRC* locale:

theorem *T1*:
  shows $stable\text{-}on\ ds\ X \longleftrightarrow (\exists\, XD\text{-}XH.\ stable\text{-}pair\text{-}on\ ds\ XD\text{-}XH \wedge X = match\ XD\text{-}XH)$
⟨*proof*⟩

end

5.2 Theorem 3: Algorithmics

Having revived Theorem 1, we reformulate *stable-pair* as a monotone (aka *isotone*) function and exploit the lattice structure of its fixed points, following Hatfield and Milgrom (2005, §II, Theorem 3). This underpins all of their results that we formulate here. Fleiner (2002, §2) provides an intuitive gloss of these definitions.

context *Contracts*
begin

definition *F1* :: *'x cfun* **where**
  $F1\ X' = -\ RH\ X'$

definition *F2* :: $'d\ set \Rightarrow 'x\ cfun$ **where**
  $F2\ ds\ X' = -\ RD\text{-}on\ ds\ X'$

definition *F* :: $'d\ set \Rightarrow 'x\ set \times 'x\ set\ dual \Rightarrow 'x\ set \times 'x\ set\ dual$ **where**
  $F\ ds = (\lambda(XD,\ XH).\ (F1\ (undual\ XH),\ dual\ (F2\ ds\ (F1\ (undual\ XH)))))$

We exploit Isabelle/HOL's ordering type classes (over the type constructors $'a\ set$ and $'a \times 'b$) to define *F*. As *F* is *antimono* (where $antimono\ f = (\forall\, x\ y.\ x \leq y \longrightarrow f\ y \leq f\ x)$ for a lattice order $\leq$) on its second argument *XH*, we adopt the dual lattice order using the type constructor $'a\ dual$, where *dual* and *undual* mediate the isomorphism on values, to satisfy Isabelle/HOL's *mono* predicate. Note we work under the *substitutes* hypothesis here.

Relating this function to *stable-pair* is syntactically awkward but straightforward:

lemma *fix-F-stable-pair-on*:
  assumes $X = F\ ds\ X$
  shows $stable\text{-}pair\text{-}on\ ds\ (map\text{-}prod\ id\ undual\ X)$

lemma *stable-pair-on-fix-F*:
  assumes $stable\text{-}pair\text{-}on\ ds\ X$
  shows $map\text{-}prod\ id\ dual\ X = F\ ds\ (map\text{-}prod\ id\ dual\ X)$

end

The function *F* is monotonic under *substitutes*.

locale $ContractsWithSubstitutes = Contracts\ +$
  assumes *Ch-substitutes*: $\forall\, h.\ substitutes\ (Ch\ h)$
begin

lemma *F1-antimono*:
  shows *antimono F1*

lemma *F2-antimono*:
  shows *antimono (F2 ds)*

lemma *F-mono*:

shows *mono* (*F ds*)

We define the extremal fixed points using Isabelle/HOL's least and greatest fixed point operators:

definition *gfp-F* :: *'b set* ⇒ *'a set* × *'a set* **where**
gfp-F ds = *map-prod id undual* (*gfp* (*F ds*))

definition *lfp-F* :: *'b set* ⇒ *'a set* × *'a set* **where**
lfp-F ds = *map-prod id undual* (*lfp* (*F ds*))

lemmas *gfp-F-stable-pair-on* = *fix-F-stable-pair-on*[*OF gfp-unfold*[*OF F-mono*], *folded gfp-F-def*]
lemmas *lfp-F-stable-pair-on* = *fix-F-stable-pair-on*[*OF lfp-unfold*[*OF F-mono*], *folded lfp-F-def*]

These last two lemmas show that the least and greatest fixed points do satisfy *stable-pair*.
Using standard fixed-point properties, we can establish:

lemma *F2-o-F1-mono*:
shows *mono* (*F2 ds* ∘ *F1*)

lemmas *F2-F1-mono* = *F2-o-F1-mono*[*unfolded o-def*]

lemma *gfp-F-lfp-F*:
shows *gfp-F ds* = (*F1* (*lfp* (*F2 ds* ∘ *F1*)), *lfp* (*F2 ds* ∘ *F1*))

end

We need hospital CFs to satisfy both *substitutes* and *irc* to relate these fixed points to stable matches.

locale *ContractsWithSubstitutesAndIRC* =
ContractsWithSubstitutes + *ContractsWithIRC*
begin

lemmas *gfp-F-stable-on* = *stable-pair-on-stable-on*[*OF gfp-F-stable-pair-on*]
lemmas *lfp-F-stable-on* = *stable-pair-on-stable-on*[*OF lfp-F-stable-pair-on*]

end

We demonstrate the effectiveness of our definitions by executing an example due to Hatfield and Milgrom (2005, p920) using Isabelle/HOL's code generator (Haftmann and Nipkow 2010). Note that, while adequate for this toy instance, the representations used here are hopelessly näive: sets are represented by lists and operations typically traverse the entire contract space. It is feasible, with more effort, to derive efficient algorithms; see, for instance, Bijlsma (1991); Bulwahn et al. (2008).

context *ContractsWithSubstitutes*
begin

lemma *gfp-F-code*[*code*]:
shows *gfp-F ds* = *map-prod id undual* (*while* (λ*A*. *F ds A* ≠ *A*) (*F ds*) *top*)

lemma *lfp-F-code*[*code*]:
shows *lfp-F ds* = *map-prod id undual* (*while* (λ*A*. *F ds A* ≠ *A*) (*F ds*) *bot*)

end

There are two hospitals and two doctors.

datatype *H2* = *H1* | *H2*

The contract space is simply the Cartesian product *D2* × *H2*.

type-synonym *X-D2-H2* = *D2* × *H2*

Doctor *D1* prefers *H1* ≻ *H2*, doctor *D2* the same *H1* ≻ *H2* (but over different contracts).

primrec *P-D2-H2-d* :: *D2* ⇒ *X-D2-H2 rel* **where**
  P-D2-H2-d D1 = *linord-of-list* [(*D1*, *H1*), (*D1*, *H2*)]
| *P-D2-H2-d D2* = *linord-of-list* [(*D2*, *H1*), (*D2*, *H2*)]

Hospital *H1* prefers $\{D1\} \succ \{D2\} \succ \emptyset$, and hospital *H2* $\{D1, D2\} \succ \{D1\} \succ \{D2\} \succ \emptyset$. We interpret these constraints as follows:

definition *P-D2-H2-H1* :: *X-D2-H2 cfun* **where**
  P-D2-H2-H1 A = (*if* (*D1*, *H1*) ∈ *A then* {(*D1*, *H1*)} *else if* (*D2*, *H1*) ∈ *A then* {(*D2*, *H1*)} *else* {})

definition *P-D2-H2-H2* :: *X-D2-H2 cfun* **where**
  P-D2-H2-H2 A =
    (*if* {(*D1*, *H2*), (*D2*, *H2*)} ⊆ *A then* {(*D1*, *H2*), (*D2*, *H2*)} *else*
     if (*D1*, *H2*) ∈ *A then* {(*D1*, *H2*)} *else if* (*D2*, *H2*) ∈ *A then* {(*D2*, *H2*)} *else* {})

primrec *P-D2-H2-h* :: *H2* ⇒ *X-D2-H2 cfun* **where**
  P-D2-H2-h H1 = *P-D2-H2-H1*
| *P-D2-H2-h H2* = *P-D2-H2-H2*⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

Isabelle's code generator requires us to hoist the relevant definitions from the locale to the top-level (see the `codegen` documentation, §7.3).

global-interpretation *P920-example*:
  ContractsWithSubstitutes fst snd P-D2-H2-d P-D2-H2-h
defines *P920-example-gfp-F* = *P920-example.gfp-F*
  and *P920-example-lfp-F* = *P920-example.lfp-F*
  and *P920-example-F* = *P920-example.F*
  and *P920-example-F1* = *P920-example.F1*
  and *P920-example-F2* = *P920-example.F2*
  and *P920-example-maxR* = *P920-example.maxR*
  and *P920-example-MaxR-f* = *P920-example.MaxR-f*
  and *P920-example-Cd* = *P920-example.Cd*
  and *P920-example-CD-on* = *P920-example.CD-on*
  and *P920-example-CH* = *P920-example.CH*
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

We can now evaluate the *gfp* of *P920-example.F* (i.e., *F* specialized to the above constants) using Isabelle's `value` antiquotation or `eval` method. This yields the (*XD*, *XH*) pair:

$$(\{(D1, H1), (D1, H2), (D2, H2)\}, \{(D1, H1), (D2, H1), (D2, H2)\})$$

The stable match is therefore {(*D1*, *H1*), (*D2*, *H2*)}.
The *lfp* of *P920-example.F* is identical to the *gfp*:

$$(\{(D1, H1), (D1, H2), (D2, H2)\}, \{(D1, H1), (D2, H1), (D2, H2)\})$$

This implies that there is only one stable match in this scenario.

5.3 Theorem 4: Optimality

Hatfield and Milgrom (2005, Theorem 4) assert that the greatest fixed point *gfp-F* of *F* yields the stable match most preferred by the doctors in the following sense:

context *Contracts*
begin

definition *doctor-optimal-match* :: *'d set* ⇒ *'x set* ⇒ *bool* **where**
  doctor-optimal-match ds Y
    ⟷ (*stable-on ds Y* ∧ (∀ *X*. ∀ *x*∈*X*. *stable-on ds X* ⟶ (∃ *y* ∈ *Y*. (*x*, *y*) ∈ *Pd* (*Xd x*))))
⟨*proof*⟩

end

In a similar sense, *lfp-F* is the doctor-pessimal match.

We state a basic doctor-optimality result in terms of *stable-pair* in the *ContractsWithSubstitutes* locale for generality; we can establish *doctor-optimal-match* only under additional constraints on hospital choice functions (see §5.1.2).

context *ContractsWithSubstitutes*
begin

context
  fixes $XD\text{-}XH :: {}'a\ set \times {}'a\ set$
  fixes $ds :: {}'b\ set$
  assumes *stable-pair-on ds XD-XH*
begin

lemma *gfp-F-upperbound*:
  shows $(fst\ XD\text{-}XH,\ dual\ (snd\ XD\text{-}XH)) \leq gfp\ (F\ ds)$

lemma *XD-XH-gfp-F*:
  shows $fst\ XD\text{-}XH \subseteq fst\ (gfp\text{-}F\ ds)$
    and $snd\ (gfp\text{-}F\ ds) \subseteq snd\ XD\text{-}XH$

lemma *lfp-F-upperbound*:
  shows $lfp\ (F\ ds) \leq (fst\ XD\text{-}XH,\ dual\ (snd\ XD\text{-}XH))$

lemma *XD-XH-lfp-F*:
  shows $fst\ (lfp\text{-}F\ ds) \subseteq fst\ XD\text{-}XH$
    and $snd\ XD\text{-}XH \subseteq snd\ (lfp\text{-}F\ ds)$

We appeal to the doctors' linear preferences to show the optimality (pessimality) of *gfp-F* (*lfp-F*) for doctors.

theorem *gfp-f-doctor-optimal*:
  assumes $x \in match\ XD\text{-}XH$
  shows $\exists\, y \in match\ (gfp\text{-}F\ ds).\ (x,\ y) \in Pd\ (Xd\ x)$

theorem *lfp-f-doctor-pessimal*:
  assumes $x \in match\ (lfp\text{-}F\ ds)$
  shows $\exists\, y \in match\ XD\text{-}XH.\ (x,\ y) \in Pd\ (Xd\ x)$

end

end

theorem (**in** *ContractsWithSubstitutesAndIRC*) *gfp-F-doctor-optimal-match*:
  shows *doctor-optimal-match ds (match (gfp-F ds))*

Conversely *lfp-F* is most preferred by the hospitals in a revealed-preference sense, and *gfp-F* least preferred. These results depend on *Ch-domain* and hence the *irc* hypothesis on hospital choice functions.

context *ContractsWithSubstitutesAndIRC*
begin

theorem *lfp-f-hospital-optimal*:
  assumes *stable-pair-on ds XD-XH*
  assumes $x \in Ch\ h\ (match\ (lfp\text{-}F\ ds))$
  shows $x \in Ch\ h\ (match\ (lfp\text{-}F\ ds) \cup match\ XD\text{-}XH)$

theorem *gfp-f-hospital-pessimal*:
  assumes *stable-pair-on ds XD-XH*

assumes $x \in Ch\ h\ (match\ XD\text{-}XH)$
shows $x \in Ch\ h\ (match\ (gfp\text{-}F\ ds) \cup match\ XD\text{-}XH)$

end

The general lattice-theoretic results of e.g. Fleiner (2002) depend on the full Tarski-Knaster fixed point theorem, which is difficult to state in the present type class-based setting. (The theorem itself is available in the Isabelle/HOL distribution but requires working with less convenient machinery.)

5.4 Theorem 5 does not hold (Hatfield and Kojima 2008)

Hatfield and Milgrom (2005, Theorem 5) claim that:

> Suppose H contains at least two hospitals, which we denote by h and h'. Further suppose that Rf $(Ch\ h)$ is not isotone, that is, contracts are not *substitutes* for h. Then there exist preference orderings for the doctors in set D, a preference ordering for a hospital h' with a single job opening such that, regardless of the preferences of the other hospitals, no stable set of contracts exists.

Hatfield and Kojima (2008, Observation 1) show this is not true: there can be stable matches even if hospital choice functions violate *substitutes*. This motivates looking for conditions weaker than *substitutes* that still guarantee stable matches, a project taken up by Hatfield and Kojima (2010); see §6. We omit their counterexample to this incorrect claim.

5.5 Theorem 6: "Vacancy chain" dynamics

Hatfield and Milgrom (2005, II(C), p923) propose a model for updating a stable match X when a doctor d' retires. Intuitively the contracts mentioning d' are discarded and a modified algorithm run from the *XH-largest* and *XD-smallest* sets determined from X. The result is another stable match where the remaining doctors $ds - \{d'\}$ are (weakly) better off and the hospitals (weakly) worse off than they were in the initial state. The proofs are essentially the same as for optimality (§5.3).

context *ContractsWithSubstitutesAndIRC*
begin

context
fixes $X :: 'a\ set$
fixes $d' :: 'b$
fixes $ds :: 'b\ set$
begin

Hatfield and Milgrom do not motivate why the process uses this functional and not F.

definition $F' :: 'a\ set \times 'a\ set\ dual \Rightarrow 'a\ set \times 'a\ set\ dual$ **where**
$F' = (\lambda(XD, XH).\ (-\ RH\ (undual\ XH),\ dual\ (-\ RD\text{-}on\ (ds{-}\{d'\})\ XD)))$

lemma F'*-apply*:
$F'\ (XD, XH) = (-\ RH\ (undual\ XH),\ dual\ (-\ RD\text{-}on\ (ds - \{d'\})\ XD))$
$\langle proof \rangle$

lemma F'*-mono*:
shows $mono\ F'$

lemma *fix-*F'*-stable-pair-on*:
$stable\text{-}pair\text{-}on\ (ds - \{d'\})\ (map\text{-}prod\ id\ undual\ A)$
if $A = F'\ A$

We model their update process using the *while* combinator, as we cannot connect it to the extremal fixed points as we did in §5.2 because we begin computing from the stable match X.

definition F'*-iter* $:: 'a\ set \times 'a\ set\ dual$ **where**

F'-*iter* = (*while* (λA. F' $A \neq A$) F' (*XD-smallest ds X*, *dual* (*XH-largest ds X*)))

abbreviation F'-*iter-match* :: *'a set* **where**
F'-*iter-match* ≡ *match* (*map-prod id undual* F'-*iter*)

context
assumes *stable-on ds X*
begin

lemma *F-start*:
shows *F ds* (*XD-smallest ds X*, *dual* (*XH-largest ds X*)) = (*XD-smallest ds X*, *dual* (*XH-largest ds X*))

lemma F'-*start*:
shows (*XD-smallest ds X*, *dual* (*XH-largest ds X*)) ≤ F' (*XD-smallest ds X*, *dual* (*XH-largest ds X*))

lemma
shows F'-*iter-stable-pair-on*: *stable-pair-on* (*ds*−{*d'*}) (*map-prod id undual* F'-*iter*) (**is** *?thesis1*)
and F'-*start-le*-F'-*iter*: (*XD-smallest ds X*, *dual* (*XH-largest ds X*)) ≤ F'-*iter* (**is** *?thesis2*)

lemma F'-*iter-match-stable-on*:
shows *stable-on* (*ds*−{*d'*}) F'-*iter-match*

theorem F'-*iter-match-doctors-weakly-better-off*:
assumes *x* ∈ *Cd d X*
assumes *d* ≠ *d'*
shows ∃ *y* ∈ *Cd d* F'-*iter-match*. (*x*, *y*) ∈ *Pd d*

theorem F'-*iter-match-hospitals-weakly-worse-off*:
assumes *x* ∈ *Ch h X*
shows *x* ∈ *Ch h* (F'-*iter-match* ∪ *X*)

Hatfield and Milgrom observe but do not prove that F'-*iter-match* is not necessarily doctor-optimal wrt the new set of doctors, even if *X* was.

These results seem incomplete. One might expect that the process of reacting to a doctor's retirement would involve considering new entrants to the workforce and allowing the set of possible contracts to be refined. There are also the questions of hospitals opening and closing.

end

end

end

5.6 Theorems 8 and 9: A "rural hospitals" theorem

Given that some hospitals are less desirable than others, the question arises of whether there is a mechanism that can redistribute doctors to under-resourced hospitals while retaining the stability of the match. Roth's *rural hospitals theorem* (Roth and Sotomayor 1990, Theorem 5.12) resolves this in the negative by showing that each doctor and hospital signs the same number of contracts in every stable match. In the context of contracts the theorem relies on the further hypothesis that hospital choices satisfy the law of aggregate demand (§4.3).

locale *ContractsWithLAD* = *Contracts* +
assumes *Ch-lad*: ∀ *h*. *lad* (*Ch h*)

locale *ContractsWithSubstitutesAndLAD* =
ContractsWithSubstitutes + *ContractsWithLAD*

We can use results that hold under *irc* by discharging that hypothesis against *lad* using the *lad-on-substitutes-on-irc-on* lemma. This is the effect of the following *sublocale* command:

sublocale *ContractsWithSubstitutesAndLAD* < *ContractsWithSubstitutesAndIRC*
⟨*proof*⟩

context *ContractsWithSubstitutesAndLAD*
begin

The following results lead to Hatfield and Milgrom (2005, Theorem 8), and the proofs go as they say. Again we state these with respect to an arbitrary solution to *stable-pair*.

context
  fixes *XD-XH* :: *'a set* × *'a set*
  fixes *ds* :: *'b set*
  assumes *stable-pair-on ds XD-XH*
begin

lemma *Cd-XD-gfp-F-card*:
  assumes *d* ∈ *ds*
  shows *card* (*Cd d* (*fst XD-XH*)) ≤ *card* (*Cd d* (*fst* (*gfp-F ds*)))

lemma *Ch-gfp-F-XH-card*:
  shows *card* (*Ch h* (*snd* (*gfp-F ds*))) ≤ *card* (*Ch h* (*snd XD-XH*))

theorem *Theorem-8*:
  shows *d* ∈ *ds* ⟹ *card* (*Cd d* (*fst XD-XH*)) = *card* (*Cd d* (*fst* (*gfp-F ds*)))
    and *card* (*Ch h* (*snd XD-XH*)) = *card* (*Ch h* (*snd* (*gfp-F ds*)))

end

Their result may be more easily understood when phrased in terms of arbitrary stable matches:

corollary *rural-hospitals-theorem*:
  assumes *stable-on ds X*
  assumes *stable-on ds Y*
  shows *d* ∈ *ds* ⟹ *card* (*Cd d X*) = *card* (*Cd d Y*)
    and *card* (*Ch h X*) = *card* (*Ch h Y*)

end

Hatfield and Milgrom (2005, Theorem 9) show that without *lad*, the rural hospitals theorem does not hold. Their proof does not seem to justify the theorem as stated (for instance, the contracts x', y' and z' need not exist), and so we instead simply provide a counterexample (discovered by `nitpick`) to the same effect.

lemma (**in** *ContractsWithSubstitutesAndIRC*) *Theorem-9-counterexample*:
  assumes *stable-on ds Y*
  assumes *stable-on ds Z*
  shows *card* (*Ch h Y*) = *card* (*Ch h Z*)
⟨*proof*⟩

datatype *X3* = *Xd1* | *Xd1'* | *Xd2*⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
primrec *X3d* :: *X3* ⇒ *D2* **where**
  X3d Xd1 = *D1*
| *X3d Xd1'* = *D1*
| *X3d Xd2* = *D2*

abbreviation *X3h* :: *X3* ⇒ *H1* **where**
  X3h - ≡ *H*

primrec *PX3d* :: *D2* ⇒ *X3 rel* **where**
  PX3d D1 = *linord-of-list* [*Xd1*, *Xd1'*]
| *PX3d D2* = *linord-of-list* [*Xd2*]

function $CX3h :: H1 \Rightarrow X3\ set \Rightarrow X3\ set$ **where**
  $CX3h\ \text{-}\ \{Xd1\} = \{Xd1\}$
| $CX3h\ \text{-}\ \{Xd1'\} = \{Xd1'\}$
| $CX3h\ \text{-}\ \{Xd2\} = \{Xd2\}$
| $CX3h\ \text{-}\ \{Xd1, Xd1'\} = \{Xd1'\}$
| $CX3h\ \text{-}\ \{Xd1, Xd2\} = \{Xd1, Xd2\}$
| $CX3h\ \text{-}\ \{Xd1', Xd2\} = \{Xd1'\}$
| $CX3h\ \text{-}\ \{Xd1, Xd1', Xd2\} = \{Xd1'\}$
| $CX3h\ \text{-}\ \{\} = \{\}$
⟨proof⟩⟨proof⟩⟨proof⟩⟨proof⟩⟨proof⟩⟨proof⟩⟨proof⟩
interpretation *Theorem-9*: *ContractsWithSubstitutesAndIRC X3d X3h PX3d CX3h*

lemma *Theorem-9-stable-Xd1'*:
  shows *Theorem-9.stable-on UNIV* $\{Xd1'\}$

lemma *Theorem-9-stable-Xd1-Xd2*:
  shows *Theorem-9.stable-on UNIV* $\{Xd1, Xd2\}$

This violates the rural hospitals theorem:

theorem
  shows *card* (*Theorem-9.CH* $\{Xd1'\}$) $\neq$ *card* (*Theorem-9.CH* $\{Xd1, Xd2\}$)

. . . which is attributed to the failure of the hospitals' choice functions to satisfy *lad*:

lemma *CX3h-not-lad*:
  shows $\neg$*lad* (*CX3h h*)

Ciupan et al. (2016) discuss an alternative approach to this result in a marriage market.

5.7 Theorems 15 and 16: Cumulative Offer Processes

The goal of Hatfield and Milgrom (2005, §V) is to connect this theory of contracts with matching to earlier work on auctions by the first of the authors, in particular by eliminating the *substitutes* hypothesis. They do so by defining a *cumulative offer process* (COP):

context *Contracts*
begin

definition $cop\text{-}F\text{-}HM :: 'd\ set \Rightarrow 'x\ set \times 'x\ set \Rightarrow 'x\ set \times 'x\ set$ **where**
  $cop\text{-}F\text{-}HM\ ds = (\lambda(XD, XH).\ (-\ RH\ XH,\ XH \cup CD\text{-}on\ ds\ (-\ RH\ XH)))$

Intuitively all of the doctors simultaneously offer their most preferred contracts that have yet to be rejected by the hospitals, and the hospitals choose amongst these and all that have been offered previously. Asking hospital choice functions to satisfy the *substitutes* condition effectively forces hospitals to consider only the contracts they have previously not rejected.
This definition is neither monotonic nor increasing (i.e., it is not the case that $\forall x.\ x \leq cop\text{-}F\text{-}HM\ ds\ x$). We rectify this by focusing on the second part of the definition.

definition $cop\text{-}F :: 'd\ set \Rightarrow 'x\ set \Rightarrow 'x\ set$ **where**
  $cop\text{-}F\ ds\ XH = XH \cup CD\text{-}on\ ds\ (-\ RH\ XH)$

lemma *cop-F-HM-cop-F*:
  shows $cop\text{-}F\text{-}HM\ ds\ XD\text{-}XH = (-\ RH\ (snd\ XD\text{-}XH),\ cop\text{-}F\ ds\ (snd\ XD\text{-}XH))$
⟨proof⟩

lemma *cop-F-increasing*:
  shows $x \leq cop\text{-}F\ ds\ x$

We have the following straightforward case distinction principles:

lemma *cop-F-cases*:
  assumes $x \in$ *cop-F ds fp*
  obtains (*fp*) $x \in$ *fp* | (*CD-on*) $x \in$ *CD-on ds* ($-$*RH fp*) $-$ *fp*
⟨*proof*⟩

lemma *CH-cop-F-cases*:
  assumes $x \in$ *CH* (*cop-F ds fp*)
  obtains (*CH*) $x \in$ *CH fp* | (*RH-fp*) $x \in$ *RH fp* | (*CD-on*) $x \in$ *CD-on ds* ($-$*RH fp*) $-$ *fp*
⟨*proof*⟩

The existence of fixed points for our earlier definitions (§5.2) was guaranteed by the Tarski-Knaster theorem, which relies on the monotonicity of the defining functional. As *cop-F* lacks this property, we appeal instead to the Bourbaki-Witt theorem for increasing functions.

interpretation *COP*: *bourbaki-witt-fixpoint Sup* $\{(x, y).\ x \leq y\}$ *cop-F ds* **for** *ds*

definition *fp-cop-F* :: *'d set* $\Rightarrow$ *'x set* **where**
  fp-cop-F ds = *COP.fixp-above ds* {}

abbreviation *cop ds* $\equiv$ *CH* (*fp-cop-F ds*)

Given that the set of contracts is finite, we avoid continuity and admissibility issues; we have the following straightforward induction principle:

lemma *fp-cop-F-induct*[*case-names base step*]:
  assumes *P* {}
  assumes $\bigwedge$*fp*. *P fp* $\Longrightarrow$ *P* (*cop-F ds fp*)
  shows *P* (*fp-cop-F ds*)

An alternative is to use the *while* combinator, which is equivalent to the above by *COP.fixp-above-conv-while*.

In any case, invariant reasoning is essential to verifying the properties of the COP, no matter how we phrase it. We develop a small program logic to ease the reuse of the invariants we prove.

definition
  valid :: *'d set* $\Rightarrow$ (*'d set* $\Rightarrow$ *'x set* $\Rightarrow$ *bool*) $\Rightarrow$ (*'d set* $\Rightarrow$ *'x set* $\Rightarrow$ *bool*) $\Rightarrow$ *bool*
where
  valid ds P Q = (*Q ds* {} $\wedge$ ($\forall$*fp*. *P ds fp* $\wedge$ *Q ds fp* $\longrightarrow$ *Q ds* (*cop-F ds fp*)))

abbreviation
  invariant :: *'d set* $\Rightarrow$ (*'d set* $\Rightarrow$ *'x set* $\Rightarrow$ *bool*) $\Rightarrow$ *bool*
where
  invariant ds P $\equiv$ *valid ds* (λ- -. *True*) *P*

Intuitively *valid ds P Q* asserts that the COP satisfies *Q* assuming that it satisfies *P*. This allows us to decompose our invariant proofs. By setting the precondition to *True*, *invariant ds P* captures the proof obligations of *fp-cop-F-induct* exactly.

The following lemmas ease the syntactic manipulation of these facts.

lemma *validI*[*case-names base step*]:
  assumes *Q ds* {}
  assumes $\bigwedge$*fp*. ⟦*P ds fp*; *Q ds fp*⟧ $\Longrightarrow$ *Q ds* (*cop-F ds fp*)
  shows *valid ds P Q*

lemma *invariant-cop-FD*:
  assumes *invariant ds P*
  assumes *P ds fp*
  shows *P ds* (*cop-F ds fp*)

lemma *invariantD*:
  assumes *invariant ds P*
  shows *P ds* (*fp-cop-F ds*)

lemma *valid-pre*:
  assumes *valid ds P′ Q*
  assumes $\bigwedge$*fp*. *P ds fp* $\Longrightarrow$ *P′ ds fp*
  shows *valid ds P Q*

lemma *valid-invariant*:
  assumes *valid ds P Q*
  assumes *invariant ds P*
  shows *invariant ds* (λ *ds fp*. *P ds fp* ∧ *Q ds fp*)

lemma *valid-conj*:
  assumes *valid ds* (λ*ds fp*. *R ds fp* ∧ *P ds fp* ∧ *Q ds fp*) *P*
  assumes *valid ds* (λ*ds fp*. *R ds fp* ∧ *P ds fp* ∧ *Q ds fp*) *Q*
  shows *valid ds R* (λ *ds fp*. *P ds fp* ∧ *Q ds fp*)

end

Hatfield and Milgrom (2005, Theorem 15) assert that *fp-cop-F* is equivalent to the doctor-offering algorithm *gfp-F*, assuming *substitutes*. (Note that the fixed points generated by increasing functions do not necessarily form a lattice, so there is not necessarily a hospital-optimal match, and indeed in general these do not exist.) Our proof is eased by the decomposition lemma *gfp-F-lfp-F* and the standard properties of fixed points in a lattice.

context *ContractsWithSubstitutes*
begin

lemma *lfp-F2-o-F1-fp-cop-F*:
  shows *lfp* (*F2 ds* ∘ *F1*) = *fp-cop-F ds*
⟨*proof*⟩

theorem *Theorem-15*:
  shows *gfp-F ds* = (− *RH* (*fp-cop-F ds*), *fp-cop-F ds*)
⟨*proof*⟩

theorem *Theorem-15-match*:
  shows *match* (*gfp-F ds*) = *CH* (*fp-cop-F ds*)
⟨*proof*⟩

end

With some auxiliary definitions, we can evaluate the COP on the example from §5.2.

lemma *P920-example-fp-cop-F-value*:
  shows *P920-example-CH* (*P920-example-fp-cop-F UNIV*) = {(*D1*, *H1*), (*D2*, *H2*)}
⟨*proof*⟩

Hatfield and Milgrom (2005, Theorem 16) assert that this process yields a stable match when we have a single hospital (now called an auctioneer) with unrestricted preferences. As before, this holds provided the auctioneer's preferences satisfy *irc*.

We begin by establishing two obvious invariants of the COP that hold in general.

context *Contracts*
begin

definition *cop-F-range-inv* :: *′d set* ⇒ *′x set* ⇒ *bool* **where**
  cop-F-range-inv ds fp ⟷ (∀ *x*∈*fp*. *x* ∈ *Field* (*Pd* (*Xd x*)) ∧ *Xd x* ∈ *ds*)

definition *cop-F-closed-inv* :: *'d set* ⇒ *'x set* ⇒ *bool* **where**
cop-F-closed-inv ds fp ⟷ (∀ *x*∈*fp*. *above* (*Pd* (*Xd x*)) *x* ⊆ *fp*)

The first, *cop-F-range-inv*, simply states that the result of the COP respects the structural conditions for doctors. The second *cop-F-closed-inv* states that the COP is upwards-closed with respect to the doctors' preferences.

lemma *cop-F-range-inv*:
shows *invariant ds cop-F-range-inv*
⟨*proof*⟩

lemma *cop-F-closed-inv*:
shows *invariant ds cop-F-closed-inv*
⟨*proof*⟩

lemmas *fp-cop-F-range-inv* = *invariantD*[*OF cop-F-range-inv*]
lemmas *fp-cop-F-range-inv′* = *fp-cop-F-range-inv*[*unfolded cop-F-range-inv-def*, *rule-format*]
lemmas *fp-cop-F-closed-inv* = *invariantD*[*OF cop-F-closed-inv*]
lemmas *fp-cop-F-closed-inv′* = *subsetD*[*OF bspec*[*OF invariantD*[*OF cop-F-closed-inv*, *unfolded cop-F-closed-inv-def*, *simplified*]]]

The only challenge in showing that the COP yields a stable match is in establishing *stable-no-blocking-on*. Our key lemma states that that if *CH* rejects all contracts for doctor *d* in *fp-cop-F*, then all contracts for *d* are in *fp-cop-F*.

lemma *cop-F-RH*:
assumes *d* ∈ *ds*
assumes *x* ∈ *Field* (*Pd d*)
assumes *aboveS* (*Pd d*) *x* ⊆ *RH fp*
shows *x* ∈ *cop-F ds fp*

lemma *fp-cop-F-all*:
assumes *d* ∈ *ds*
assumes *d* ∉ *Xd* ‘ *CH* (*fp-cop-F ds*)
shows *Field* (*Pd d*) ⊆ *fp-cop-F ds*

Aygün and Sönmez (2012b) observe that any blocking contract must be weakly preferred by its doctor to anything in the outcome of the *fp-cop-F*:

lemma *fp-cop-F-preferred*:
assumes *y* ∈ *CD-on ds* (*CH* (*fp-cop-F ds*) ∪ *X″*)
assumes *x* ∈ *CH* (*fp-cop-F ds*)
assumes *Xd x* = *Xd y*
shows (*x*, *y*) ∈ *Pd* (*Xd x*)

The headline lemma cobbles these results together.

lemma *X″-closed*:
assumes *X″* ⊆ *CD-on ds* (*CH* (*fp-cop-F ds*) ∪ *X″*)
shows *X″* ⊆ *fp-cop-F ds*
⟨*proof*⟩

The *irc* constraint on the auctioneer's preferences is needed for *stable-no-blocking* and their part of *individually-rational*.

end

context *ContractsWithIRC*
begin

lemma *cop-stable-no-blocking-on*:
shows *stable-no-blocking-on ds* (*cop ds*)
⟨*proof*⟩

theorem *Theorem-16*:
  assumes h: $(UNIV::'c\ set) = \{h\}$
  shows *stable-on ds* (*cop ds*) (**is** *stable-on ds ?fp*)
⟨*proof*⟩

end

5.8 Concluding remarks

From Hatfield and Milgrom (2005), we have not shown Theorems 2, 7, 13 and 14, all of which are intended to position their results against prior work in this space. We delay establishing their strategic results (Theorems 10, 11 and 12) to §8, after we have developed more useful invariants for the COP.

By assuming *irc*, Aygün and Sönmez (2012b) are essentially trading on Plott's path independence condition (§4.7), as observed by Chambers and Yenmez (2013). The latter show that these results generalize naturally to many-to-many matches, where doctors also use path-independent choice functions; see also Fleiner (2003).

For many applications, however, *substitutes* proves to be too strong a condition. The COP of §5.7 provides a way forward, as we discuss in the next section.

6 Hatfield and Kojima (2010): Substitutes and stability for matching with contracts

Hatfield and Kojima (2010) set about weakening *substitutes* and therefore making the *cumulative offer processes* (COPs, §5.7) applicable to more matching problems. In doing so they lose the lattice structure of the stable matches, which necessitates redeveloping the results of §5.

In contrast to the COP of §5.7, Hatfield and Kojima (2010) develop and analyze a *single-offer* variant, where only one doctor (who has no held contract) proposes per round. The order of doctors making offers is not specified. We persist with the simultaneous-offer COP as it is deterministic. See Hirata and Kasuya (2014) for equivalence arguments.

We begin with some observations due to Aygün and Sönmez. Firstly, as for the matching-with-contracts setting of §5, Aygün and Sönmez (2012a) demonstrate that these results depend on hospital preferences satisfying *irc*. We do not formalize their examples. Secondly, an alternative to hospitals having choice functions (as we have up to now) is for the hospitals to have preference orders over sets, which is suggested by both Hatfield and Milgrom (2005) (weakly) and Hatfield and Kojima (2010). Aygün and Sönmez (2012a, §2) argue that this approach is under-specified and propose to define *Ch* as choosing amongst maximal elements of some non-strict preference order (i.e., including indifference). They then claim that this is equivalent to taking *Ch* as primitive, and so we continue down that path.

6.1 Theorem 1: the COP yields a stable match under *bilateral substitutes*

The weakest replacement condition suggested by Hatfield and Kojima (2010, §1) for the *substitutes* condition on hospital choice functions is termed *bilateral substitutes*:

> Contracts are *bilateral substitutes* for a hospital if there are no two contracts x and z and a set of contracts Y with other doctors than those associated with x and z such that the hospital that regards Y as available wants to sign z if and only if x becomes available. In other words, contracts are bilateral substitutes when any hospital, presented with an offer from a doctor he does not currently employ, never wishes to also hire another doctor he does not currently employ at a contract he previously rejected.

Note that this constraint is specific to this matching-with-contracts setting, unlike those of §4.

context *Contracts*
begin

definition *bilateral-substitutes-on* :: $'x\ set \Rightarrow\ 'x\ cfun \Rightarrow bool$ **where**
  bilateral-substitutes-on A f

$\longleftrightarrow \neg(\exists\, B{\subseteq}A.\ \exists\, a\ b.\ \{a,\ b\} \subseteq A \wedge Xd\ a \notin Xd\ `\ B \wedge Xd\ b \notin Xd\ `\ B$
$\wedge\ b \notin f\ (B \cup \{b\}) \wedge b \in f\ (B \cup \{a,\ b\}))$

abbreviation *bilateral-substitutes* :: *'x cfun* ⇒ *bool* **where**
bilateral-substitutes ≡ *bilateral-substitutes-on UNIV*

lemma *bilateral-substitutes-on-def2*:
bilateral-substitutes-on A f
$\longleftrightarrow (\forall\, B{\subseteq}A.\ \forall\, a{\in}A.\ \forall\, b{\in}A.\ Xd\ a \notin Xd\ `\ B \wedge Xd\ b \notin Xd\ `\ B \wedge b \notin f\ (B \cup \{b\}) \longrightarrow b \notin f\ (B \cup \{a,\ b\}))$

lemma *substitutes-on-bilateral-substitutes-on*:
assumes *substitutes-on A f*
shows *bilateral-substitutes-on A f*

Aygün and Sönmez (2012a, §4, Definition 5) give the following equivalent definition:

lemma *bilateral-substitutes-on-def3*:
bilateral-substitutes-on A f
$\longleftrightarrow (\forall\, B{\subseteq}A.\ \forall\, a{\in}A.\ \forall\, b{\in}A.\ b \notin f\ (B \cup \{b\}) \wedge b \in f\ (B \cup \{a,\ b\}) \longrightarrow Xd\ a \in Xd\ `\ B \vee Xd\ b \in Xd\ `\ B)$

end

As before, we define a series of locales that capture the relevant hypotheses about hospital choice functions.

locale *ContractsWithBilateralSubstitutes* = *Contracts* +
assumes *Ch-bilateral-substitutes*: ∀ *h*. *bilateral-substitutes* (*Ch h*)

sublocale *ContractsWithSubstitutes* < *ContractsWithBilateralSubstitutes*

locale *ContractsWithBilateralSubstitutesAndIRC* =
ContractsWithBilateralSubstitutes + *ContractsWithIRC*

sublocale *ContractsWithSubstitutesAndIRC* < *ContractsWithBilateralSubstitutesAndIRC*

context *ContractsWithBilateralSubstitutesAndIRC*
begin

The key difficulty in showing the stability of the result of the COP under this condition (Hatfield and Kojima 2010, Theorem 1) is in proving that it ensures we get an *allocation*; the remainder of the proof of §5.7 (for a single hospital, where this property is trivial) goes through unchanged. We avail ourselves of Hirata and Kasuya (2014, Lemma), which they say is a restatement of the proof of Hatfield and Kojima (2010, Theorem 1). See also Aygün and Sönmez (2012a, Appendix A).

lemma *bilateral-substitutes-lemma*:
assumes $Xd\ x \notin Xd\ `\ Ch\ h\ X$
assumes $d \notin Xd\ `\ Ch\ h\ X$
assumes $d \neq Xd\ x$
shows $d \notin Xd\ `\ Ch\ h\ (insert\ x\ X)$
⟨*proof*⟩

Our proof essentially adds the inductive details these earlier efforts skipped over. It is somewhat complicated by our use of the simultaneous-offer COP.

lemma *bilateral-substitutes-lemma-union*:
assumes $Xd\ `\ Ch\ h\ X \cap Xd\ `\ Y = \{\}$
assumes $d \notin Xd\ `\ Ch\ h\ X$
assumes $d \notin Xd\ `\ Y$
assumes *allocation Y*
shows $d \notin Xd\ `\ Ch\ h\ (X \cup Y)$

lemma *cop-F-CH-CD-on-disjoint*:

assumes *cop-F-closed-inv ds fp*
assumes *cop-F-range-inv ds fp*
shows Xd ‘ $CH\ fp \cap Xd$ ‘ $(CD\text{-}on\ ds\ (-\ RH\ fp) - fp) = \{\}$

Our key lemma shows that we effectively have *substitutes* for rejected contracts, provided the relevant doctor does not have a contract held with the relevant hospital. Note the similarity to Theorem 4 (§6.3).

lemma *cop-F-RH-mono*:
assumes *cop-F-closed-inv ds fp*
assumes *cop-F-range-inv ds fp*
assumes $Xd\ x \notin Xd$ ‘ $Ch\ (Xh\ x)\ fp$
assumes $x \in RH\ fp$
shows $x \in RH\ (cop\text{-}F\ ds\ fp)$
⟨*proof*⟩

lemma *cop-F-allocation-inv*:
valid ds (λ*ds fp. cop-F-range-inv ds fp* $\wedge$ *cop-F-closed-inv ds fp*) (λ*ds fp. allocation* (*CH fp*))
⟨*proof*⟩

lemma *fp-cop-F-allocation*:
shows *allocation* (*cop ds*)

theorem *Theorem-1*:
shows *stable-on ds* (*cop ds*)

end

Hatfield and Kojima (2010, §3.1) provide an example that shows that the traditional optimality and strategic results do not hold under *bilateral-substitutes*, which motivates looking for a stronger condition that remains weaker than *substitutes*.

Their example involves two doctors, two hospitals, and five contracts.

datatype $X5 = Xd1 \mid Xd1' \mid Xd2 \mid Xd2' \mid Xd2''$

primrec $X5d :: X5 \Rightarrow D2$ **where**
$X5d\ Xd1 = D1$
$\mid X5d\ Xd1' = D1$
$\mid X5d\ Xd2 = D2$
$\mid X5d\ Xd2' = D2$
$\mid X5d\ Xd2'' = D2$

primrec $X5h :: X5 \Rightarrow H2$ **where**
$X5h\ Xd1 = H1$
$\mid X5h\ Xd1' = H1$
$\mid X5h\ Xd2 = H1$
$\mid X5h\ Xd2' = H2$
$\mid X5h\ Xd2'' = H1$

primrec $PX5d :: D2 \Rightarrow X5\ rel$ **where**
$PX5d\ D1 = linord\text{-}of\text{-}list\ [Xd1,\ Xd1']$
$\mid PX5d\ D2 = linord\text{-}of\text{-}list\ [Xd2,\ Xd2',\ Xd2'']$

primrec $CX5h :: H2 \Rightarrow X5\ cfun$ **where**
$CX5h\ H1\ A =$
$(if\ \{Xd1',\ Xd2\} \subseteq A\ then\ \{Xd1',\ Xd2\}\ else$
$if\ \{Xd2''\} \subseteq A\ then\ \{Xd2''\}\ else$
$if\ \{Xd1\} \subseteq A\ then\ \{Xd1\}\ else$
$if\ \{Xd1'\} \subseteq A\ then\ \{Xd1'\}\ else$
$if\ \{Xd2\} \subseteq A\ then\ \{Xd2\}\ else\ \{\})$

| *CX5h H2 A* = { *x* . *x* ∈ *A* ∧ *x* = *Xd2′* }
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩

interpretation *BSI*: *Contracts X5d X5h PX5d CX5h*

lemma *CX5h-bilateral-substitutes*:
  shows *BSI.bilateral-substitutes* (*CX5h h*)
⟨*proof*⟩

lemma *CX5h-irc*:
  shows *irc* (*CX5h h*)
⟨*proof*⟩

interpretation *BSI*: *ContractsWithBilateralSubstitutesAndIRC X5d X5h PX5d CX5h*

There are two stable matches in this model.

⟨*proof*⟩⟨*proof*⟩

lemma *BSI-stable*:
  shows *BSI.stable X* ⟷ *X* = {*Xd1*, *Xd2′*} ∨ *X* = {*Xd1′*, *Xd2*}⟨*proof*⟩

Therefore there is no doctor-optimal match under these preferences:

lemma
  ¬(∃ (*Y*::*X5 set*). *BSI.doctor-optimal-match UNIV Y*)
⟨*proof*⟩

6.2 Theorem 3: *pareto separability* relates *unilateral substitutes* and *substitutes*

Hatfield and Kojima (2010, §4) proceed to define *unilateral substitutes*:

> [P]references satisfy *unilateral substitutes* if whenever a hospital rejects the contract *z* when that is the only contract with *Xd z* available, it still rejects the contract *z* when the choice set expands.

context *Contracts*
begin

definition *unilateral-substitutes-on* :: *′x set* ⇒ *′x cfun* ⇒ *bool* **where**
  unilateral-substitutes-on A f
    ⟷ ¬(∃ *B*⊆*A*. ∃ *a b*. {*a*, *b*} ⊆ *A* ∧ *Xd b* ∉ *Xd* ‘ *B* ∧ *b* ∉ *f* (*B* ∪ {*b*}) ∧ *b* ∈ *f* (*B* ∪ {*a*, *b*}))

abbreviation *unilateral-substitutes* :: *′x cfun* ⇒ *bool* **where**
  unilateral-substitutes ≡ *unilateral-substitutes-on UNIV*

lemma *unilateral-substitutes-on-def2*:
  unilateral-substitutes-on A f
    ⟷ (∀ *B*⊆*A*. ∀ *a*∈*A*. ∀ *b*∈*A*. *Xd b* ∉ *Xd* ‘ *B* ∧ *b* ∉ *f* (*B* ∪ {*b*}) ⟶ *b* ∉ *f* (*B* ∪ {*a*, *b*}))

Aygün and Sönmez (2012a, §4, Definition 6) give the following equivalent definition:

lemma *unilateral-substitutes-on-def3*:
  unilateral-substitutes-on A f
    ⟷ (∀ *B*⊆*A*. ∀ *a*∈*A*. ∀ *b*∈*A*. *b* ∉ *f* (*B* ∪ {*b*}) ∧ *b* ∈ *f* (*B* ∪ {*a*, *b*}) ⟶ *Xd b* ∈ *Xd* ‘ *B*)

lemma *substitutes-on-unilateral-substitutes-on*:
  assumes *substitutes-on A f*
  shows *unilateral-substitutes-on A f*

lemma *unilateral-substitutes-on-bilateral-substitutes-on*:
  assumes *unilateral-substitutes-on A f*

shows *bilateral-substitutes-on A f*

The following defines locales for the *unilateral-substitutes* hypothesis, and inserts these between those for *substitutes* and *bilateral-substitutes*.

end

locale *ContractsWithUnilateralSubstitutes* = *Contracts* +
assumes *Ch-unilateral-substitutes*: $\forall h.$ *unilateral-substitutes* (*Ch h*)

sublocale *ContractsWithUnilateralSubstitutes* < *ContractsWithBilateralSubstitutes*

sublocale *ContractsWithSubstitutes* < *ContractsWithUnilateralSubstitutes*

locale *ContractsWithUnilateralSubstitutesAndIRC* =
ContractsWithUnilateralSubstitutes + *ContractsWithIRC*

sublocale *ContractsWithUnilateralSubstitutesAndIRC* < *ContractsWithBilateralSubstitutesAndIRC*

sublocale *ContractsWithSubstitutesAndIRC* < *ContractsWithUnilateralSubstitutesAndIRC*

Hatfield and Kojima (2010, Theorem 3) relate *unilateral-substitutes* to *substitutes* using *Pareto separability*:

> Preferences are *Pareto separable* for a hospital if the hospital's choice between x and x', two [distinct] contracts with the same doctor, does not depend on what other contracts the hospital has access to.

This result also depends on *irc*.

context *Contracts*
begin

definition *pareto-separable-on* :: $'x$ *set* $\Rightarrow$ *bool* **where**
pareto-separable-on A
$\longleftrightarrow (\forall B \subseteq A.\ \forall C \subseteq A.\ \forall a\ b.\ \{a, b\} \subseteq A \wedge a \neq b \wedge Xd\ a = Xd\ b \wedge Xh\ a = Xh\ b$
$\wedge\ a \in Ch\ (Xh\ b)\ (B \cup \{a, b\}) \longrightarrow b \notin Ch\ (Xh\ b)\ (C \cup \{a, b\}))$

abbreviation *pareto-separable* :: *bool* **where**
pareto-separable $\equiv$ *pareto-separable-on UNIV*

lemma *substitutes-on-pareto-separable-on*:
assumes $\forall h.$ *substitutes-on A* (*Ch h*)
shows *pareto-separable-on A*
⟨*proof*⟩

lemma *unilateral-substitutes-on-pareto-separable-on-substitutes-on*:
assumes $\forall h.$ *unilateral-substitutes-on A* (*Ch h*)
assumes $\forall h.$ *irc-on A* (*Ch h*)
assumes *pareto-separable-on A*
shows *substitutes-on A* (*Ch h*)
⟨*proof*⟩

theorem *Theorem-3*:
assumes $\forall h.$ *irc-on A* (*Ch h*)
shows ($\forall h.$ *substitutes-on A* (*Ch h*)) $\longleftrightarrow$ ($\forall h.$ *unilateral-substitutes-on A* (*Ch h*) $\wedge$ *pareto-separable-on A*)

end

6.2.1 Afacan and Turhan (2015): *doctor separability* relates bi- and unilateral substitutes

context *Contracts*

begin

Afacan and Turhan (2015, Theorem 1) relate *bilateral-substitutes* and *unilateral-substitutes* using *doctor separability*:

> *[Doctor separability (DS)]* says that if a doctor is not chosen from a set of contracts in the sense that no contract of him is selected, then that doctor should still not be chosen unless a contract of a new doctor (that is, doctor having no contract in the given set of contracts) becomes available. For practical purposes, we can consider DS as capturing contracts where certain groups of doctors are substitutes. [footnote: If $Xd\ x \notin Xd$ ‘ $Ch\ h\ (Y \cup \{x, z\})$, then doctor $Xd\ x$ is not chosen. And under DS, he continues not to be chosen unless a new doctor comes. Hence, we can interpret it as the doctors in the given set of contracts are substitutes.]

definition *doctor-separable-on* :: $'x$ *set* $\Rightarrow$ $'x$ *cfun* $\Rightarrow$ *bool* **where**
  doctor-separable-on A f
    $\longleftrightarrow (\forall B \subseteq A.\ \forall a\ b\ c.\ \{a, b, c\} \subseteq A \wedge Xd\ a \neq Xd\ b \wedge Xd\ b = Xd\ c \wedge Xd\ a \notin Xd$ ‘ $f\ (B \cup \{a, b\})$
      $\longrightarrow Xd\ a \notin Xd$ ‘ $f\ (B \cup \{a, b, c\}))$

abbreviation *doctor-separable* :: $'x$ *cfun* $\Rightarrow$ *bool* **where**
  doctor-separable $\equiv$ *doctor-separable-on UNIV*

lemma *unilateral-substitutes-on-doctor-separable-on*:
  assumes *unilateral-substitutes-on A f*
  assumes *irc-on A f*
  assumes $\forall B \subseteq A.$ *allocation* $(f\ B)$
  assumes *f-range-on A f*
  shows *doctor-separable-on A f*
⟨*proof*⟩

lemma *bilateral-substitutes-on-doctor-separable-on-unilateral-substitutes-on*:
  assumes *bilateral-substitutes-on A f*
  assumes *doctor-separable-on A f*
  assumes *f-range-on A f*
  shows *unilateral-substitutes-on A f*
⟨*proof*⟩

theorem *unilateral-substitutes-on-doctor-separable-on-bilateral-substitutes-on*:
  assumes *irc-on A f*
  assumes $\forall B \subseteq A.$ *allocation* $(f\ B)$ — A rephrasing of *Ch-singular*.
  assumes *f-range-on A f*
  shows *unilateral-substitutes-on A f* $\longleftrightarrow$ *bilateral-substitutes-on A f* $\wedge$ *doctor-separable-on A f*

Afacan and Turhan (2015, Remark 2) observe the independence of the *doctor-separable*, *pareto-separable* and *bilateral-substitutes* conditions.

end

6.3 Theorems 4 and 5: Doctor optimality

context *ContractsWithUnilateralSubstitutesAndIRC*
begin

We return to analyzing the COP following Hatfield and Kojima (2010). The next goal is to establish a doctor-optimality result for it in the spirit of §5.3.

We first show that, with hospital choice functions satisfying *unilateral-substitutes*, we effectively have the *substitutes* condition for all contracts that have been rejected. In other words, hospitals never renegotiate with doctors.

The proof is by induction over the finite set Y.

lemma
  assumes $Xd\ x \notin Xd$ ‘ $Ch\ h\ X$

assumes $x \in X$
shows *no-renegotiation-union*: $x \notin Ch\ h\ (X \cup Y)$
and $x \notin Ch\ h\ (insert\ x\ ((X \cup Y) - \{z.\ Xd\ z = Xd\ x\}))$

To discharge the first antecedent of this lemma, we need an invariant for the COP that asserts that, for each doctor d, there is a subset of the contracts currently offered by d that was previously uniformly rejected by the COP, for each contract that is rejected at the current step. To support a later theorem (see §6.3) we require these subsets to be upwards-closed with respect to the doctor's preferences.

definition
$cop\text{-}F\text{-}rejected\text{-}inv :: 'b\ set \Rightarrow 'a\ set \Rightarrow bool$
where
$cop\text{-}F\text{-}rejected\text{-}inv\ ds\ fp \longleftrightarrow (\forall x{\in}RH\ fp.\ \exists fp'{\subseteq}fp.\ x \in fp' \wedge above\ (Pd\ (Xd\ x))\ x \subseteq fp' \wedge Xd\ x \notin Xd\ \text{‘}\ CH\ fp')$

lemma *cop-F-rejected-inv*:
shows $valid\ ds\ (\lambda ds\ fp.\ cop\text{-}F\text{-}range\text{-}inv\ ds\ fp \wedge cop\text{-}F\text{-}closed\text{-}inv\ ds\ fp \wedge allocation\ (CH\ fp))\ cop\text{-}F\text{-}rejected\text{-}inv$

lemma *fp-cop-F-rejected-inv*:
shows $cop\text{-}F\text{-}rejected\text{-}inv\ ds\ (fp\text{-}cop\text{-}F\ ds)$

Hatfield and Kojima (2010, Theorem 4) assert that we effectively recover *substitutes* for the contracts relevant to the COP. We cannot adopt their phrasing as it talks about the execution traces of the COP, and not just its final state. Instead we present the result we use, which relates two consecutive states in an execution trace of the COP:

theorem *Theorem-4*:
assumes $cop\text{-}F\text{-}rejected\text{-}inv\ ds\ fp$
assumes $x \in RH\ fp$
shows $x \in RH\ (cop\text{-}F\ ds\ fp)$

Another way to interpret *cop-F-rejected-inv* is to observe that the doctor-optimal match contains the least preferred of the contracts that the doctors have offered.

corollary *fp-cop-F-worst*:
assumes $x \in cop\ ds$
assumes $y \in fp\text{-}cop\text{-}F\ ds$
assumes $Xd\ y = Xd\ x$
shows $(x,\ y) \in Pd\ (Xd\ x)$

The doctor optimality result, Theorem 5, hinges on showing that no contract in any stable match is ever rejected.

definition
$theorem\text{-}5\text{-}inv :: 'b\ set \Rightarrow 'a\ set \Rightarrow bool$
where
$theorem\text{-}5\text{-}inv\ ds\ fp \longleftrightarrow RH\ fp \cap \bigcup\{X.\ stable\text{-}on\ ds\ X\} = \{\}$
⟨*proof*⟩
lemma *theorem-5-inv*:
shows $valid\ ds\ (\lambda ds\ fp.\ cop\text{-}F\text{-}range\text{-}inv\ ds\ fp \wedge cop\text{-}F\text{-}closed\text{-}inv\ ds\ fp$
$\wedge\ allocation\ (CH\ fp) \wedge cop\text{-}F\text{-}rejected\text{-}inv\ ds\ fp)\ theorem\text{-}5\text{-}inv$
⟨*proof*⟩

lemma *fp-cop-F-theorem-5-inv*:
shows $theorem\text{-}5\text{-}inv\ ds\ (fp\text{-}cop\text{-}F\ ds)$

theorem *Theorem-5*:
assumes $stable\text{-}on\ ds\ X$
assumes $x \in X$
shows $\exists y \in cop\ ds.\ (x,\ y) \in Pd\ (Xd\ x)$
⟨*proof*⟩

theorem *fp-cop-F-doctor-optimal-match*:
  shows *doctor-optimal-match ds* (*cop ds*)

end

The next lemma demonstrates the *opposition of interests* of doctors and hospitals: if all doctors weakly prefer one stable match to another, then the hospitals weakly prefer the converse.

As we do not have linear preferences for hospitals, we use revealed preference and hence assume *irc* holds of hospital choice functions. Our definition of the doctor-preferred ordering *dpref* follows the Isabelle/HOL convention of putting the larger (more preferred) element on the right, and takes care with unemployment.

context *Contracts*
begin

definition *dpref* :: *'x set* ⇒ *'x set* ⇒ *bool* **where**
  dpref X Y = (∀ *x*∈*X*. ∃ *y*∈*Y*. (*x*, *y*) ∈ *Pd* (*Xd x*))

end

context *ContractsWithIRC*
begin

theorem *Lemma-1*:
  assumes *stable-on ds Y*
  assumes *stable-on ds Z*
  assumes *dpref Z Y*
  assumes *x* ∈ *Ch h Z*
  shows *x* ∈ *Ch h* (*Y* ∪ *Z*)
⟨*proof*⟩

end

Hatfield and Kojima (2010, Corollary 1 (of Theorem 5 and Lemma 1)): *unilateral-substitutes* implies there is a hospital-pessimal match, which is indeed the doctor-optimal one.

context *ContractsWithUnilateralSubstitutesAndIRC*
begin

theorem *Corollary-1*:
  assumes *stable-on ds Z*
  shows *dpref Z* (*cop ds*)
    and *x* ∈ *Z* ⟹ *x* ∈ *Ch* (*Xh x*) (*cop ds* ∪ *Z*)
⟨*proof*⟩

Hatfield and Kojima (2010, p1717) show that there is not always a hospital-optimal/doctor-pessimal match when hospital preferences satisfy *unilateral-substitutes*, in contrast to the situation under *substitutes* (see §5.3). This reflects the loss of the lattice structure.

end

6.4 Theorem 6: A "rural hospitals" theorem

Hatfield and Kojima (2010, Theorem 6) demonstrates a "rural hospitals" theorem for the COP assuming hospital choice functions satisfy *unilateral-substitutes* and *lad*, as for §5.6. However Aygün and Sönmez (2012a, §4, Example 1) observe that *lad-on-substitutes-on-irc-on* does not hold with *bilateral-substitutes* instead of *substitutes*, and their Example 3 similarly for *unilateral-substitutes*. Moreover *fp-cop-F* can yield an unstable allocation with just these two hypotheses. Ergo we need to assume *irc* even when we have *lad*, unlike before (see §5.6).

This theorem is the foundation for all later strategic results.

locale *ContractsWithUnilateralSubstitutesAndIRCAndLAD* = *ContractsWithUnilateralSubstitutesAndIRC* + *ContractsWithLAD*

sublocale *ContractsWithSubstitutesAndLAD* < *ContractsWithUnilateralSubstitutesAndIRCAndLAD*

context *ContractsWithUnilateralSubstitutesAndIRCAndLAD*
begin

context
  fixes *ds* :: *'b set*
  fixes *X* :: *'a set*
  assumes *stable-on ds X*
begin

The proofs of these first two lemmas are provided by Hatfield and Kojima (2010, Theorem 6). We treat unemployment in the definition of the function A as we did in §5.1.3.

lemma *RHT-Cd-card*:
  assumes $d \in ds$
  shows *card* (*Cd d X*) $\leq$ *card* (*Cd d* (*cop ds*))

lemma *RHT-Ch-card*:
  shows *card* (*Ch h* (*fp-cop-F ds*)) $\leq$ *card* (*Ch h X*)
⟨*proof*⟩

The top-level proof is the same as in §5.6.

lemma *Theorem-6-fp-cop-F*:
  shows $d \in ds \Longrightarrow$ *card* (*Cd d X*) = *card* (*Cd d* (*cop ds*))
    and *card* (*Ch h X*) = *card* (*Ch h* (*fp-cop-F ds*))
⟨*proof*⟩

end

theorem *Theorem-6*:
  assumes *stable-on ds X*
  assumes *stable-on ds Y*
  shows $d \in ds \Longrightarrow$ *card* (*Cd d X*) = *card* (*Cd d Y*)
    and *card* (*Ch h X*) = *card* (*Ch h Y*)

end

6.5 Concluding remarks

We next discuss a kind of interference between doctors termed *bossiness* in §7. This has some implications for the strategic issues we discuss in §8.

7 Kojima (2010): The non-existence of a stable and non-bossy mechanism

Kojima (2010) says that "a mechanism is *nonbossy* if an agent cannot change [the] allocation of other agents unless doing so also changes her own allocation." He shows that no mechanism can be both *stable-on* and *nonbossy* in a one-to-one marriage market. We establish this result in our matching-with-contracts setting here.

There are two complications. Firstly, as not all agent preferences yield stable matches (unlike the marriage market), we constrain hospital choice functions to satisfy *ContractsWithBilateralSubstitutesAndIRC*, which is the weakest condition formalized here that ensures that *fp-cop-F* yields stable matches. (We note that it is not the weakest condition guaranteeing the existence of stable matches.)

Secondly, non-bossiness needs to separately treat the preferences of the doctors and the choice functions of the hospitals.

We work in the *Contracts* locale for its types and the constants *Xd* and *Xh*. To account for the quantification over preferences, we directly use some raw constants from the *Contracts* locale.

context *Contracts*
begin

abbreviation (*input*) $\mathit{mechanism\text{-}domain} :: ('d \Rightarrow 'x\ \mathit{rel}) \Rightarrow ('h \Rightarrow 'x\ \mathit{cfun}) \Rightarrow \mathit{bool}$ **where**
$\mathit{mechanism\text{-}domain} \equiv \mathit{ContractsWithBilateralSubstitutesAndIRC}\ Xd\ Xh$

definition $\mathit{nonbossy} :: 'd\ \mathit{set} \Rightarrow ('d, 'h, 'x)\ \mathit{mechanism} \Rightarrow \mathit{bool}$ **where**
$\mathit{nonbossy}\ ds\ \varphi \longleftrightarrow$
$(\forall\, Pd\ Pd'\ Ch.\ \forall\, d{\in}ds.\ \mathit{mechanism\text{-}domain}\ Pd\ Ch \wedge \mathit{mechanism\text{-}domain}\ (Pd(d{:=}Pd'))\ Ch \longrightarrow$
$dX\ (\varphi\ Pd\ Ch\ ds)\ d = dX\ (\varphi\ (Pd(d{:=}Pd'))\ Ch\ ds)\ d \longrightarrow \varphi\ Pd\ Ch\ ds = \varphi\ (Pd(d{:=}Pd'))\ Ch\ ds)$
$\wedge\ (\forall\, Pd\ Ch\ Ch'\ h.\ \mathit{mechanism\text{-}domain}\ Pd\ Ch \wedge \mathit{mechanism\text{-}domain}\ Pd\ (Ch(h{:=}Ch')) \longrightarrow$
$hX\ (\varphi\ Pd\ Ch\ ds)\ h = hX\ (\varphi\ Pd\ (Ch(h{:=}Ch'))\ ds)\ h \longrightarrow \varphi\ Pd\ Ch\ ds = \varphi\ Pd\ (Ch(h{:=}Ch'))\ ds)$

definition $\mathit{mechanism\text{-}stable} :: 'd\ \mathit{set} \Rightarrow ('d, 'h, 'x)\ \mathit{mechanism} \Rightarrow \mathit{bool}$ **where**
$\mathit{mechanism\text{-}stable}\ ds\ \varphi$
$\longleftrightarrow (\forall\, Pd\ Ch.\ \mathit{mechanism\text{-}domain}\ Pd\ Ch \longrightarrow \mathit{Contracts.stable\text{-}on}\ Pd\ Ch\ ds\ (\varphi\ Pd\ Ch\ ds))$
⟨*proof*⟩⟨*proof*⟩
end

The proof is somewhat similar to those for Roth's impossibility results (see, for instance, Roth and Sotomayor (1990, Theorem 4.4)). It relies on the existence of at least three doctors, three hospitals, and a complete set of contracts between these. The following locale captures a suitable set of constraints.

locale *BossyConstants* =
fixes $Xd :: 'x \Rightarrow 'd$
fixes $Xh :: 'x \Rightarrow 'h$
fixes $d1h1\ d1h2\ d1h3 :: 'x$
fixes $d2h1\ d2h2\ d2h3 :: 'x$
fixes $d3h1\ d3h2\ d3h3 :: 'x$
fixes $ds :: 'd\ \mathit{set}$
assumes ds: $\mathit{distinct}\ [Xd\ d1h1,\ Xd\ d2h1,\ Xd\ d3h1]$
assumes hs: $\mathit{distinct}\ [Xh\ d1h1,\ Xh\ d1h2,\ Xh\ d1h3]$
assumes $Xd\text{-}xs$:
$Xd\ `\ \{d1h2,\ d1h3\} = \{Xd\ d1h1\}$
$Xd\ `\ \{d2h2,\ d2h3\} = \{Xd\ d2h1\}$
$Xd\ `\ \{d3h2,\ d3h3\} = \{Xd\ d3h1\}$
assumes $Xh\text{-}xs$:
$Xh\ `\ \{d2h1,\ d3h1\} = \{Xh\ d1h1\}$
$Xh\ `\ \{d2h2,\ d3h2\} = \{Xh\ d1h2\}$
$Xh\ `\ \{d2h3,\ d3h3\} = \{Xh\ d1h3\}$
assumes dset: $\{Xd\ d1h1,\ Xd\ d2h1,\ Xd\ d3h1\} \subseteq ds$

locale *ContractsWithBossyConstants* =
Contracts + *BossyConstants*
begin

abbreviation (*input*) $d1 \equiv Xd\ d1h1$
abbreviation (*input*) $d2 \equiv Xd\ d2h1$
abbreviation (*input*) $d3 \equiv Xd\ d3h1$

abbreviation (*input*) $h1 \equiv Xh\ d1h1$
abbreviation (*input*) $h2 \equiv Xh\ d1h2$
abbreviation (*input*) $h3 \equiv Xh\ d1h3$
⟨*proof*⟩

We proceed to show that variations on the following preferences for doctors and hospitals force a stable mechanism to be bossy. Recall that *linord-of-list* constructs a linear order from a list of elements greatest to least. The hospital choice functions take at most one contract from those on offer, and are again ordered from most preferable to least.

definition $BPd :: 'b \Rightarrow 'a\ \mathit{rel}$ **where**

$BPd \equiv map\text{-}of\text{-}default\ \{\}\ [\ (d1,\ linord\text{-}of\text{-}list\ [d1h3,\ d1h2,\ d1h1])$
$\quad ,\ (d2,\ linord\text{-}of\text{-}list\ [d2h3,\ d2h2,\ d2h1])$
$\quad ,\ (d3,\ linord\text{-}of\text{-}list\ [d3h1,\ d3h2,\ d3h3])\ ]$

abbreviation $mkhord :: 'd\ list \Rightarrow 'd\ cfun$ **where**
$mkhord\ xs\ X \equiv set\text{-}option\ (List.find\ (\lambda x.\ x{\in}X)\ xs)$

definition $BCh :: 'c \Rightarrow 'a\ cfun$ **where**
$BCh \equiv map\text{-}of\text{-}default\ (\lambda\text{-}.\ \{\})\ [\ (h1,\ mkhord\ [d1h1,\ d2h1,\ d3h1])$
$\quad ,\ (h2,\ mkhord\ [])$
$\quad ,\ (h3,\ mkhord\ [d3h3,\ d2h3,\ d1h3])\ ]$

Interpreting the *Contracts* locale gives us access to some useful constants.

interpretation *Bossy*: *Contracts Xd Xh BPd BCh*

lemma *BPd-BCh-mechanism-domain*:
shows *mechanism-domain BPd BCh*
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
lemma *Bossy-stable*:
shows $Bossy.stable\text{-}on\ ds\ X \longleftrightarrow X = \{d1h1,\ d3h3\}$⟨*proof*⟩

The second preference order has doctor *d2* reject all contracts and is otherwise the same as the first.

definition $BPd' :: 'b \Rightarrow 'a\ rel$ **where**
$BPd' = BPd(d2 := \{\})$

interpretation *Bossy'*: *Contracts Xd Xh BPd' BCh*

lemma *BPd'-BCh-mechanism-domain*:
shows *mechanism-domain BPd' BCh*
⟨*proof*⟩⟨*proof*⟩
lemma *Bossy'-stable*:
shows $Bossy'.stable\text{-}on\ ds\ X \longleftrightarrow X = \{d1h3,\ d3h1\} \lor X = \{d1h1,\ d3h3\}$⟨*proof*⟩

The third preference order adjusts the choice function of hospital *h2* and is otherwise the same as the second.

definition $BCh' :: 'c \Rightarrow 'a\ cfun$ **where**
$BCh' \equiv BCh(h2 := mkhord\ [d1h2,\ d2h2,\ d3h2])$

interpretation *Bossy''*: *Contracts Xd Xh BPd' BCh'*

lemma *BPd'-BCh'-mechanism-domain*:
shows *mechanism-domain BPd' BCh'*
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
lemma *Bossy''-stable*:
shows $Bossy''.stable\text{-}on\ ds\ X \longleftrightarrow X = \{d3h1,\ d1h3\}$⟨*proof*⟩

theorem *Theorem-1*:
shows $\neg(mechanism\text{-}stable\ ds\ \varphi \land nonbossy\ ds\ \varphi)$
⟨*proof*⟩

In particular, the COP (see §6) is bossy as it always yields stable matches under *mechanism-stable*.

theorem *Theorem-1-COP*:
$\neg nonbossy\ ds\ Contracts.cop$
⟨*proof*⟩

end

Therefore doctors can interfere with other doctors' allocations under the COP without necessarily disadvantaging themselves, which has implications for the notion of *group strategy-proof* (Hatfield and Kojima 2009); see §8.2.

8 Strategic results

We proceed to establish a series of strategic results for the COP (see §5.7 and §6), making use of the invariants we developed for it. These results also apply to the matching-with-contracts setting of §5, and where possible we specialize our lemmas to it.

8.1 Hatfield and Milgrom (2005): Theorems 10 and 11: Truthful revelation as a Dominant Strategy

Theorems 10 and 11 demonstrate that doctors cannot obtain better results for themselves in the doctor-optimal match (i.e., *cop ds*, equal to *match* (*gfp-F ds*) by *Theorem-15-match* assuming hospital preferences satisfy *substitutes*) by misreporting their preferences. (See Roth and Sotomayor (1990, §4.2) for a discussion about the impossibility of a mechanism being strategy-proof for all agents.)

Hatfield and Milgrom (2005, §III(B)) provide the following intuition:

> We will show the positive incentive result for the doctor-offering algorithm in two steps which highlight the different roles of the two preference assumptions. First, we show that the *substitutes* condition, by itself, guarantees that doctors cannot benefit by exaggerating the ranking of an unattainable contract. More precisely, if there exists a preferences list for a doctor d such that d obtains contract x by submitting this list, then d can also obtain x by submitting a preference list that includes only contract x [Theorem 10]. Second, we will show that adding the law of aggregate demand guarantees that a doctor does at least as well as reporting truthfully as by reporting any singleton [Theorem 11]. Together, these are the dominant strategy result.

We prove Theorem 10 via a lemma that states that the contracts above $x \in X$ for some stable match X with respect to manipulated preferences Pd $(Xd\ x)$ do not improve the outcome for doctor $Xd\ x$ with respect to their true preferences Pd' $(Xd\ x)$ in the doctor-optimal match for Pd'.

This is weaker than Hatfield and Kojima (2009, Lemma 1) (see §8.2) as we do not guarantee that the allocation does not change. By the bossiness result of §7, such manipulations can change the outcomes of the other doctors; this lemma establishes that only weak improvements are possible.

context *ContractsWithUnilateralSubstitutesAndIRC*
begin

context
  fixes $d' :: 'b$
  fixes $Pd' :: 'b \Rightarrow 'a\ rel$
  assumes Pd'-d'-*linear*: *Linear-order* $(Pd'\ d')$
  assumes Pd'-d'-*range*: *Field* $(Pd'\ d') \subseteq \{y.\ Xd\ y = d'\}$
  assumes Pd': $\forall\, d.\ d \neq d' \longrightarrow Pd'\ d = Pd\ d$
begin
⟨*proof*⟩⟨*proof*⟩
interpretation *PdXXX*: *ContractsWithUnilateralSubstitutesAndIRC* $Xd\ Xh\ Pd'\ Ch$

theorem *Pd-above-irrelevant*:
  assumes d'-*Field*: $dX\ X\ d' \subseteq$ *Field* $(Pd'\ d')$
  assumes d'-*Above*: *Above* $(Pd'\ d')$ $(dX\ X\ d') \subseteq$ *Above* $(Pd\ d')$ $(dX\ X\ d')$
  assumes $x \in X$
  assumes *stable-on ds X*
  shows $\exists\, y \in PdXXX.cop\ ds.\ (x,\ y) \in Pd'\ (Xd\ x)$
⟨*proof*⟩

end

end

We now specialize this lemma to Theorem 10 by defining a preference order for the doctors where distinguished doctors *ds* submit single preferences for the contracts they receive in the doctor-optimal match.

The function *override-on* $f\ g\ A = (\lambda a.\ if\ a \in A\ then\ g\ a\ else\ f\ a)$ denotes function update at several points.

context *Contracts*
begin

definition *Pd-singletons-for-ds* :: $'x$ *set* $\Rightarrow$ $'d$ *set* $\Rightarrow$ $'d \Rightarrow$ $'x$ *rel* **where**
Pd-singletons-for-ds X ds $\equiv$ *override-on Pd* $(\lambda d.\ dX\ X\ d \times dX\ X\ d)$ *ds*
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
end

We interpret our *ContractsWithUnilateralSubstitutesAndIRC* locale with respect to this updated preference order, which gives us the stable match and properties of it.

context *ContractsWithUnilateralSubstitutesAndIRC*
begin

context
fixes *ds* :: $'b$ *set*
fixes *X* :: $'a$ *set*
assumes *stable-on ds X*
begin

interpretation
Singleton-for-d: *ContractsWithUnilateralSubstitutesAndIRC Xd Xh Pd-singletons-for-ds X* $\{d\}$ *Ch* **for** *d*

Our version of Hatfield and Milgrom (2005, Theorem 10) (for the COP) states that if a doctor submits a preference order containing just x, where x is their contract in some stable match X, then that doctor receives exactly x in the doctor-optimal match and all other doctors do at least as well.

theorem *Theorem-10-fp-cop-F*:
assumes $x \in X$
shows $\exists\, y \in$ *Singleton-for-d.cop d ds*. $(x, y) \in$ *Pd-singletons-for-ds X* $\{d\}$ $(Xd\ x)$
⟨*proof*⟩

end

end

We can recover the original Theorem 10 by specializing this result to *gfp-F*.

context *ContractsWithSubstitutesAndIRC*
begin

interpretation
Singleton-for-d: *ContractsWithSubstitutesAndIRC Xd Xh Pd-singletons-for-ds* (*match* (*gfp-F ds*)) $\{d\}$ *Ch*
for *ds d*

theorem *Theorem-10*:
assumes $x \in$ *match* (*gfp-F ds*)
shows $\exists\, y \in$ *match* (*Singleton-for-d.gfp-F ds d ds*). $(x, y) \in$ *Pd-singletons-for-ds* (*match* (*gfp-F ds*)) $\{d\}$ $(Xd\ x)$
⟨*proof*⟩

corollary *Theorem-10-d*:
assumes $x \in$ *match* (*gfp-F ds*)
shows $x \in$ *match* (*Singleton-for-d.gfp-F ds* $(Xd\ x)$ *ds*)
⟨*proof*⟩

end

The second theorem (Hatfield and Milgrom 2005, Theorem 11) depends on both Theorem 10 and the rural hospitals theorem (§5.6, §6.4). It shows that, assuming everything else is fixed, if doctor d' obtains contract x with (manipulated) preferences $Pd\ d'$ in the doctor-optimal match, then they will obtain a contract at least as good by submitting their true preferences $Pd'\ d'$ (with respect to these true preferences).

locale *TruePrefs* = *Contracts* +
  fixes $x :: 'a$
  fixes $X :: 'a\ set$
  fixes $ds :: 'b\ set$
  fixes $Pd' :: 'b \Rightarrow 'a\ rel$
  assumes x: $x \in X$
  assumes X: *stable-on ds X*
  assumes Pd'-d'-x: $x \in Field\ (Pd'\ (Xd\ x))$
  assumes Pd'-d'-*linear*: *Linear-order* $(Pd'\ (Xd\ x))$
  assumes Pd'-d'-*range*: $Field\ (Pd'\ (Xd\ x)) \subseteq \{y.\ Xd\ y = Xd\ x\}$
  assumes Pd': $\forall\ d.\ d \neq Xd\ x \longrightarrow Pd'\ d = Pd\ d$
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
locale *ContractsWithUnilateralSubstitutesAndIRCAndLADAndTruePrefs* =
  ContractsWithUnilateralSubstitutesAndIRCAndLAD + *TruePrefs*
begin

interpretation *TruePref*: *ContractsWithUnilateralSubstitutesAndIRCAndLAD Xd Xh* Pd' *Ch*

interpretation *TruePref-tax*: *ContractsWithUnilateralSubstitutesAndIRCAndLAD Xd Xh* Pd'*-tax Ch*

interpretation
  Singleton-for-d: *ContractsWithUnilateralSubstitutesAndIRCAndLAD Xd Xh Pd-singletons-for-ds X* $\{Xd\ x\}$ *Ch*
⟨*proof*⟩
lemma *Theorem-11-*Pd'*-tax*:
  shows $\exists\ y \in TruePref\text{-}tax.cop\ ds.\ (x,\ y) \in Pd'\text{-}tax\ (Xd\ x)$
⟨*proof*⟩

theorem *Theorem-11-fp-cop-F*:
  shows $\exists\ y \in TruePref.cop\ ds.\ (x,\ y) \in Pd'\ (Xd\ x)$
⟨*proof*⟩

end

locale *ContractsWithSubstitutesAndLADAndTruePrefs* =
  ContractsWithSubstitutesAndLAD + *TruePrefs*

sublocale *ContractsWithSubstitutesAndLADAndTruePrefs*
    < *ContractsWithUnilateralSubstitutesAndIRCAndLADAndTruePrefs*

context *ContractsWithSubstitutesAndLADAndTruePrefs*
begin

interpretation *TruePref*: *ContractsWithSubstitutesAndLAD Xd Xh* Pd' *Ch*

theorem *Theorem-11*:
  shows $\exists\ y \in match\ (TruePref.gfp\text{-}F\ ds).\ (x,\ y) \in Pd'\ (Xd\ x)$
⟨*proof*⟩

end

Note that this theorem depends on the hypotheses introduced by the *TruePrefs* locale, and only applies to doctor *Xd x*. The following sections show more general and syntactically self-contained results.

We omit Hatfield and Milgrom (2005, Theorem 12), which demonstrates the almost-necessity of LAD for truth revelation to be the dominant strategy for doctors.

8.2 Hatfield and Kojima (2009, 2010): The doctor-optimal match is group strategy-proof

Hatfield and Kojima (2010, Theorem 7) assert that the COP is group strategy-proof, which we define below. We begin by focusing on a single agent (Hatfield and Kojima 2009):

> A mechanism φ is *strategy-proof* if, for any preference profile Pd, there is no doctor d and preferences Pd' such that d strictly prefers y_d to x_d according to $Pd\ d$, where x_d and y_d are the (possibly null) contracts for d in $\varphi\ Pd$ and $\varphi\ Pd(d := Pd')$, respectively.

The syntax $f(a := b) = (\lambda x.\ \mathit{if}\ x = a\ \mathit{then}\ b\ \mathit{else}\ f\ x)$ denotes function update at a point.
We make this definition in the *Contracts* locale to avail ourselves of some types and the *Xd* and *Xh* constants. We also restrict hospital preferences to those that guarantee our earlier strategic results. As *gfp-F* requires these to satisfy the stronger *substitutes* constraint for stable matches to exist, we now deal purely with the COP.

context *Contracts*
begin

abbreviation (*input*) *mechanism-domain* :: $('d \Rightarrow 'x\ rel) \Rightarrow ('h \Rightarrow 'x\ cfun) \Rightarrow bool$ **where**
  mechanism-domain $\equiv$ *ContractsWithUnilateralSubstitutesAndIRCAndLAD Xd Xh*

definition *strategy-proof* :: $'d\ set \Rightarrow ('d, 'h, 'x)\ mechanism \Rightarrow bool$ **where**
  strategy-proof ds $\varphi \longleftrightarrow$
   $(\forall\, Pd\ Ch.\ \mathit{mechanism\text{-}domain}\ Pd\ Ch \longrightarrow$
   $\neg(\exists\, d{\in}ds.\ \exists\, Pd'.\ \mathit{mechanism\text{-}domain}\ (Pd(d{:=}Pd'))\ Ch$
    $\wedge\ (\exists\, y{\in}\varphi\ (Pd(d{:=}Pd'))\ Ch\ ds.\ y \in \mathit{AboveS}\ (Pd\ d)\ (dX\ (\varphi\ Pd\ Ch\ ds)\ d))))$
⟨*proof*⟩

theorem *fp-cop-F-strategy-proof*:
  shows *strategy-proof ds Contracts.cop* (**is** *strategy-proof - ?*φ)

end

The adaptation to groups is straightforward (Hatfield and Kojima 2009, 2010):

> A mechanism φ is *group strategy-proof* if, for any preference profile Pd, there is no group of doctors $ds' \subseteq ds$ and a preference profile Pd' such that every $d \in ds'$ strictly prefers y_d to x_d according to $Pd\ d$, where x_d and y_d are the (possibly null) contracts for d in $\varphi\ Pd$ and $\varphi\ Pd(d_1 := Pd'\ d_1, \ldots, d_n := Pd'\ d_n)$, respectively.

This definition requires all doctors in the coalition to strictly prefer the outcome with manipulated preferences, as Kojima's bossiness results (see §7) show that a doctor may influence other doctors' allocations without affecting their own. See Hatfield and Kojima (2009, §3) for discussion, and also Roth and Sotomayor (1990, Chapter 4); in particular their §4.3.1 discusses the robustness of these results and exogenous transfers.

context *Contracts*
begin

definition *group-strategy-proof* :: $'d\ set \Rightarrow ('d, 'h, 'x)\ mechanism \Rightarrow bool$ **where**
  group-strategy-proof ds $\varphi \longleftrightarrow$
   $(\forall\, Pd\ Ch.\ \mathit{mechanism\text{-}domain}\ Pd\ Ch \longrightarrow$
   $\neg(\exists\, ds'{\subseteq}ds.\ ds' \neq \{\} \wedge (\exists\, Pd'.\ \mathit{mechanism\text{-}domain}\ (\mathit{override\text{-}on}\ Pd\ Pd'\ ds')\ Ch$
    $\wedge\ (\forall\, d{\in}ds'.\ \exists\, y{\in}\varphi\ (\mathit{override\text{-}on}\ Pd\ Pd'\ ds')\ Ch\ ds.\ y \in \mathit{AboveS}\ (Pd\ d)\ (dX\ (\varphi\ Pd\ Ch\ ds)\ d)))))$
⟨*proof*⟩
lemma *group-strategy-proof-strategy-proof*:
  assumes *group-strategy-proof ds* φ
  shows *strategy-proof ds* φ

end

Perhaps surprisingly, Hatfield and Kojima (2010, Lemma 1, for a single doctor) assert that shuffling any contract above the doctor-optimal one to the top of a doctor's preference order preserves exactly the doctor-optimal match, which on the face of it seems to contradict the bossiness result of §7: by the earlier strategy-proofness results, this cannot affect the outcome for that particular doctor, but by bossiness it may affect others. The key observation is that this manipulation preserves blocking coalitions in the presence of *lad*.

This result is central to showing the group-strategy-proofness of the COP.

context *Contracts*
begin

definition *shuffle-to-top* :: $'x$ *set* $\Rightarrow$ $'d$ $\Rightarrow$ $'x$ *rel* **where**
  $\mathit{shuffle\text{-}to\text{-}top}\ Y = (\lambda d.\ Pd\ d - dX\ Y\ d \times \mathit{UNIV} \cup (\mathit{Domain}\ (Pd\ d) \cup dX\ Y\ d) \times dX\ Y\ d)$

definition *Pd-shuffle-to-top* :: $'d$ *set* $\Rightarrow$ $'x$ *set* $\Rightarrow$ $'d$ $\Rightarrow$ $'x$ *rel* **where**
  $\mathit{Pd\text{-}shuffle\text{-}to\text{-}top}\ ds'\ Y = \mathit{override\text{-}on}\ Pd\ (\mathit{shuffle\text{-}to\text{-}top}\ Y)\ ds'$
⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩
end

context *ContractsWithUnilateralSubstitutesAndIRCAndLAD*
begin

lemma *Lemma-1*:
  assumes *allocation Y*
  assumes *III*: $\forall\, d{\in}ds''.\ \exists\, y{\in}Y.\ y \in \mathit{AboveS}\ (Pd\ d)\ (dX\ (\mathit{cop}\ ds)\ d)$
  shows $\mathit{cop}\ ds = \mathit{Contracts.cop}\ (\mathit{Pd\text{-}shuffle\text{-}to\text{-}top}\ ds''\ Y)\ \mathit{Ch}\ ds$
⟨*proof*⟩
  ⟨*proof*⟩
  ⟨*proof*⟩
    ⟨*proof*⟩

The top-level theorem states that the COP is group strategy proof. To account for the quantification over preferences, we directly use the raw constants from the *Contracts* locale.

theorem *fp-cop-F-group-strategy-proof*:
  shows *group-strategy-proof ds Contracts.cop*
    (**is** *group-strategy-proof* - $?\varphi$)
⟨*proof*⟩

end

Again, this result does not directly apply to *gfp-F* due to the mechanism domain hypothesis.

Finally, Hatfield and Kojima (2010, Corollary 2) (respectively, Hatfield and Kojima (2009, Corollary 1)) assert that the COP (*gfp-F*) is "weakly Pareto optimal", i.e., that there is no *individually-rational* allocation that every doctor strictly prefers to the doctor-optimal match.

context *ContractsWithUnilateralSubstitutesAndIRCAndLAD*
begin

theorem *Corollary-2*:
  assumes $ds \neq \{\}$
  shows $\neg(\exists\, Y.\ \mathit{individually\text{-}rational\text{-}on}\ ds\ Y$
    $\wedge\ (\forall\, d{\in}ds.\ \exists\, y{\in}Y.\ y \in \mathit{AboveS}\ (Pd\ d)\ (dX\ (\mathit{cop}\ ds)\ d)))$
⟨*proof*⟩

end

Roth and Sotomayor (1990, §4.4) discuss how the non-proposing agents can strategise to improve their outcomes in one-to-one matches.

9 Concluding remarks

We conclude with a brief and inexhaustive survey of related work.

9.1 Related work

Computer-assisted and formal reasoning. Bijlsma (1991) gives a formal pencil-and-paper derivation of the Gale-Shapley deferred-acceptance algorithm under total strict preferences and one-to-one matching (colloquially, a marriage market). He provides termination and complexity arguments, and discusses representation issues. Hamid and Castleberry (2010) treat the same algorithm in the Coq proof assistant, give a termination proof and show that it always yields a stable match. Both focus more on reasoning about programs than the theory of stable matches. Intriguingly, the latter claims that Akamai uses (modified) stable matching to assign clients to servers in their content distribution network.

Brandt and Geist (2014) use SAT technology to find results in social choice theory. They claim that the encodings used by general purpose tools like `nitpick` are too inefficient for their application.

Stable matching. In addition to the monographs Gusfield and Irving (1989); Manlove (2013); Roth and Sotomayor (1990), Roth (2008) provides a good overview up to 2007 of open problems and other aspects of this topic that we did not explore here. Sönmez and Switzer (2013) incorporate quotas and put the COP to work at the United States Military Academy. Andersson and Ehlers (2016) analyze the possibility of matching of refugees with landlords in Sweden (without mentioning matching with contracts).

One of the more famous applications of matching theory is to kidney donation (Roth 2015), a *repugnant market* where the economists' basic tool of pricing things is considered verboten. These markets are sometimes, but not always, two-sided – kidneys are often exchanged due to compatibility issues, but there are also altruistic donations and recipients who cannot reciprocate – and so the model we discussed here is not applicable. Instead generalizations of Gale's *top trading cycles* algorithm are pressed into service (Abdulkadirolu and Sönmez 1999; Shapley and Scarf 1974; Sönmez and Ünver 2010). Much recent work has hybridized these approaches – for instance, Dworczak (2016) uses a combination to enumerate all stable matches.

Echenique (2012) shows that the matching with contracts model of §5 is no more general than that of Kelso and Crawford (1982) (a job matching market with salaries). Schlegel generalizes this result to the COP setting of §6, and moreover shows how lattice structure can be recovered there, which yields a hospital-proposing deferred-acceptance algorithm that relies only on unilaterally substitutable hospital choice functions. See Hatfield and Kominers (2016) for a discussion of the many-to-many case.

Roth and Sotomayor (1990, Theorem 2.33) point to alternatives to the deferred-acceptance algorithm, and to more general matching scenarios involving couples and roommates. Manlove (2013) provides a comprehensive survey of matching with preferences.

Further results: COP. Afacan (2014) explores the following two properties:

> *[Population monotonicity]* says that no doctor is to be worse off whenever some others leave the market. *[Resource monotonicity]*, on the other hand, requires that no doctor should lose whenever hospitals start hiring more doctors.

He shows that the COP is population and resource monotonic under *irc* and *bilateral_substitutes*. Also Afacan (2015) characterizes the COP by the properties *truncation proof* ("no doctor can ever benefit from truncating his preferences") and *invariant to lower tail preferences change* ("any doctor's assignment does not depend on his preferences over worse contracts"); that the COP satisfies these properties was demonstrated in §6. See also Hatfield et al. (2016) for another set of conditions that characterize the COP.

Hirata and Kasuya (2016) show how the strategic results can be obtained without the rural hospitals theorem, in a setting that requires *irc* but not substitutability.

Further results: Strategy. There are many different ways to think about the manipulation of economic mechanisms. Some continue in the game-theoretic tradition (Gonczarowski 2014), and, for instance, compare the manipulability of mechanisms that yield stable matches (Chen et al. 2016). Techniques from computer science help refine the notion of strategy-proofness (Ashlagi and Gonczarowski 2015) and enable complexity-theoretic arguments (Aziz et al. 2015; Deng et al. 2016). Kojima and Pathak (2009) have analyzed the scope for manipulation in large matching markets.

10 Acknowledgements

I thank Rob Irving for a copy of his excellent monograph (Gusfield and Irving 1989), Jasmin C. Blanchette for help with nitpick, Andreas Lochbihler for his Bourbaki-Witt Fixpoint theory, Orhan Aygün for a helpful discussion of Aygün and Sönmez (2012a), and Roman Werpachowski for general comments.

References

A. Abdulkadirolu and T. Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88(2):233 – 260, 1999. doi: 10.1006/jeth.1999.2553.

M. O. Afacan. Some further properties of the cumulative offer process. Unpublished, August 2014. URL http://ssrn.com/abstract=2478418.

M. O. Afacan. Characterizations of the cumulative offer process. Unpublished, August 2015. URL http://ssrn.com/abstract=2653751.

M. O. Afacan and B. Turhan. On relationships between substitutes conditions. *Economics Letters*, 126:10 – 12, 2015. doi: 10.1016/j.econlet.2014.11.007.

M. Aizerman and A. Malishevski. General theory of best variants choice: Some aspects. *IEEE Transactions on Automatic Control*, 26(5):1030–1040, October 1981. doi: 10.1109/TAC.1981.1102777.

T. Andersson and L. Ehlers. Assigning refugees to landlords in Sweden: Stable maximum matchings. Working Paper 2016:18, Department of Economics, School of Economics and Management, Lund University, July 2016. URL http://project.nek.lu.se/publications/workpap/papers/wp16_18.pdf.

I. Ashlagi and Y. A. Gonczarowski. No stable matching mechanism is obviously strategy-proof. *CoRR*, abs/1511.00452, 2015. URL http://arxiv.org/abs/1511.00452.

O. Aygün and T. Sönmez. The importance of irrelevance of rejected contracts in matching under weakened substitutes conditions. Working Paper, June 2012a.

O. Aygün and T. Sönmez. Matching with contracts: The critical role of irrelevance of rejected contracts. Working Paper. Abridged in AER 103(5) as "Matching with Contracts: Comment" in 2013., May 2012b.

H. Aziz, H. G. Seedig, and J. K. von Wedel. On the susceptibility of the deferred acceptance algorithm. In G. Weiss, P. Yolum, R. H. Bordini, and E. Elkind, editors, *AAMAS 2015*, pages 939–947. ACM, 2015.

L. Bijlsma. Formal derivation of a stable marriage algorithm. In W.H.J. Feijen and A.J.M van Gastreren, editors, *C.S. Scholten dedicata: van oude machines en nieuwe rekenwijzen.* Academic Service Schoonhoven, 1991.

J. C. Blanchette, C. Kaliszyk, L. C. Paulson, and J. Urban. Hammering towards QED. *J. Formalized Reasoning*, 9 (1):101–148, 2016. doi: 10.6092/issn.1972-5787/4593.

K. C. Border. Introductory notes on preference and rational choice, April 2012. URL http://people.hss.caltech.edu/~kcb/Notes.shtml.

F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. In A. L. C. Bazzan, M. N. Huhns, A. Lomuscio, and P. Scerri, editors, *AAMAS '14*, pages 1193–1200. IFAAMAS/ACM, 2014.

L. Bulwahn, A. Krauss, F. Haftmann, L. Erkök, and J. Matthews. Imperative functional programming with Isabelle/HOL. In O. A. Mohamed, C. Muñoz, and S. Tahar, editors, *TPHOLs '08*, volume 5170 of *LNCS*, pages 352–367. Springer, 2008.

M. B. Caminati, M. Kerber, C. Lange, and C. Rowat. VCG — combinatorial Vickrey-Clarke-Groves auctions. *Archive of Formal Proofs*, 2015, 2015a. URL https://isa-afp.org/entries/Vickrey_Clarke_Groves.shtml.

M. B. Caminati, M. Kerber, C. Lange, and C. Rowat. Sound auction specification and implementation. In T. Roughgarden, M. Feldman, and M. Schwarz, editors, *EC '15*, pages 547–564, New York, NY, USA, 2015b. ACM. doi: 10.1145/2764468.2764511.

C. P. Chambers and M. B. Yenmez. Choice and matching. Under review, July 2013. URL http://econweb.ucsd.edu/~cpchambers/papers.html.

P. Chen, M. Egesdal, M. Pycia, and M. B. Yenmez. Manipulability of stable mechanisms. *American Economic Journal: Microeconomics*, 8(2):202–14, May 2016. doi: 10.1257/mic.20150035.

A. Ciupan, J. W. Hatfield, and S. D. Kominers. An elementary proof of the lone wolf theorem. Unpublished, August 2016.

Y. Deng, W. Shen, and P. Tang. Coalition manipulations of the gale-shapley algorithm. *CoRR*, abs/1502.07823, 2016. URL http://arxiv.org/abs/1502.07823.

P. Dworczak. Deferred acceptance with compensation chains. In V. Conitzer, D. Bergemann, and Y. Chen, editors, *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 65–66. ACM, 2016. doi: 10.1145/2940716.2940727.

F. Echenique. Contracts versus salaries in matching. *American Economic Review*, 102(1):594–601, 2012. doi: 10.1257/aer.102.1.594.

T. Fleiner. *Stable and Crossing Structures*. PhD thesis, Technische Universiteit Eindhoven, 2000.

T. Fleiner. Some results on stable matchings and fixed points. Technical Report tr-2002-08, Egerváry Research Group on Combinatorial Optimization, Budapest, Hungary, March 2002. URL http://www.cs.elte.hu/egres/www/tr-02-08.html.

T. Fleiner. A fixed-point approach to stable matchings and some applications. *Mathematics of Operations Research*, 28(1):103–126, 2003. doi: 10.1287/moor.28.1.103.14256. URL http://www.cs.elte.hu/egres/www/tr-01-01.html.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, January 1962.

Y. A. Gonczarowski. Manipulation of stable matchings using minimal blacklists. In M. Babaioff, V. Conitzer, and D. Easley, editors, *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*, page 449. ACM, 2014. doi: 10.1145/2600057.2602840.

D. Gusfield and R. W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge, MA, USA, 1989.

F. Haftmann and T. Nipkow. Code generation via higher-order rewrite systems. In M. Blume, N. Kobayashi, and G. Vidal, editors, *FLOPS*, volume 6009 of *LNCS*. Springer, 2010.

N. A. Hamid and C. Castleberry. Formally certified stable marriages. In H. C. Cunningham, P. Ruth, and N. A. Kraft, editors, *Proceedings of the 48th Annual Southeast Regional Conference, 2010, Oxford, MS, USA, April 15-17, 2010*, page 34. ACM, 2010. doi: 10.1145/1900008.1900056.

S. O. Hansson and T. Grüne-Yanoff. Preferences. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford, winter 2012 edition, 2012. URL http://plato.stanford.edu/archives/win2012/entries/preferences/.

J. W. Hatfield and F. Kojima. Matching with contracts: Comment. *American Economic Review*, 98(3):1189–94, 2008. doi: 10.1257/aer.98.3.1189.

J. W. Hatfield and F. Kojima. Group incentive compatibility for matching with contracts. *Games and Economic Behavior*, 67(2):745–749, 2009. doi: 10.1016/j.geb.2009.01.007.

J. W. Hatfield and F. Kojima. Substitutes and stability for matching with contracts. *Journal of Economic Theory*, 145(5):1704–1723, 2010. doi: 10.1016/j.jet.2010.01.007.

J. W. Hatfield and P. R. Milgrom. Matching with contracts. *American Economic Review*, 95(4):913–935, 2005. doi: 10.1257/0002828054825466.

J. W. Hatfield, S. D. Kominers, and A. Westkamp. Stability, strategy-proofness, and cumulative offer mechanisms. presented at MATCHUP 2015, July 2016.

J.W. Hatfield and S.D. Kominers. Contract design and stability in many-to-many matching. *Games and Economic Behavior*, 2016. doi: 10.1016/j.geb.2016.01.002.

D. Hirata and Y. Kasuya. Cumulative offer process is order-independent. *Economics Letters*, 124(1):37 – 40, 2014. doi: 10.1016/j.econlet.2014.04.008.

D. Hirata and Y. Kasuya. On stable and strategy-proof rules in matching markets with contracts. Unpublished, August 2016.

A. S. Kelso and V. P. Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica*, 50(6): 1483–1504, November 1982.

F. Kojima. Impossibility of stable and nonbossy matching mechanisms. *Economics Letters*, 107(1):69 – 70, 2010. doi: 10.1016/j.econlet.2009.12.029.

F. Kojima and P. A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627, 2009. doi: 10.1257/aer.99.3.608.

D. F. Manlove. *Algorithmics of Matching Under Preferences*, volume 2 of *Series on Theoretical Computer Science*. WorldScientific, 2013. doi: 10.1142/8591.

H. Moulin. Choice functions over a finite set: a summary. *Social Choice and Welfare*, 2(2):147–160, September 1985. doi: 10.1007/BF00437315. URL http://eprints.gla.ac.uk/92191/.

T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, Berlin, 2002.

C. R. Plott. Path independence, rationality and social choice. *Econometrica*, 41(6):1075–1091, 1973.

A. E. Roth. *Who gets what – and why: the new economics of matchmaking and market design*. Houghton Mifflin Harcourt, 2015.

A. E. Roth and M. A. Sotomayor. *Two-sided matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric society monographs. Cambridge University Press, Cambridge, United Kingdom, 1990.

Alvin Roth. Deferred acceptance algorithms: history, theory, practice, and open questions. *International Journal of Game Theory*, 36(3):537–569, March 2008. URL http://www.nber.org/papers/w13225.

J. C. Schlegel. Contracts versus salaries in matching: A general result. *Journal of Economic Theory*, 159, Part A: 552–573, 2015. doi: 10.1016/j.jet.2015.07.015.

A. K. Sen. *Collective Choice and Social Welfare*. Holden Day, 1970.

A. K. Sen. Choice functions and revealed preference. *The Review of Economic Studies*, 38(3):307–317, July 1971.

L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37, 1974. doi: 10.1016/0304-4068(74)90033-0.

M. Sotomayor. A non-constructive elementary proof of the existence of stable marriages. *Games and Economic Behavior*, 13(1):135 – 137, 1996. doi: 10.1006/game.1996.0029.

T. Sönmez and T. B. Switzer. Matching with (branch-of-choice) contracts at the United States Military Academy. *Econometrica*, 81(2):451–488, 2013. doi: 10.3982/ECTA10570.

T. Sönmez and M. U. Ünver. House allocation with existing tenants: A characterization. *Games and Economic Behavior*, 69(2):425 – 445, 2010. doi: 10.1016/j.geb.2009.10.010.