Symmetric Groups and Quotient Complexity of Boolean Operations*

Jason Bell[1], Janusz Brzozowski[2], Nelma Moreira[3], and Rogério Reis[3]

[1] Department of Pure Mathematics, University of Waterloo,
Waterloo, ON, Canada N2L 3G1
jpbell@uwaterloo.ca
[2] David R. Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, Canada N2L 3G1
brzozo@uwaterloo.ca
[3] CMUP & DCC, Faculdade de Ciências da Universidade do Porto,
Rua do Campo Alegre, 4169–007 Porto, Portugal
{nam,rvr}@dcc.fc.up.pt

Abstract. The quotient complexity of a regular language L is the number of left quotients of L, which is the same as the state complexity of L. Suppose that L and L' are binary regular languages with quotient complexities m and n, and that the subgroups of permutations in the transition semigroups of the minimal deterministic automata accepting L and L' are the symmetric groups S_m and S_n of degrees m and n, respectively. Denote by $\circ$ any binary boolean operation that is not a constant and not a function of one argument only. For $m, n \geq 2$ with $(m, n) \notin \{(2, 2), (3, 4), (4, 3), (4, 4)\}$ we prove that the quotient complexity of $L \circ L'$ is mn if and only either (a) $m \neq n$ or (b) $m = n$ and the bases (ordered pairs of generators) of S_m and S_n are not conjugate. For $(m, n) \in \{(2, 2), (3, 4), (4, 3), (4, 4)\}$ we give examples to show that this need not hold. In proving these results we generalize the notion of uniform minimality to direct products of automata. We also establish a non-trivial connection between complexity of boolean operations and group theory.

Keywords: Boolean operation, quotient complexity, regular language, state complexity, symmetric group, transition semigroup.

1 Motivation

The *left quotient,* or simply *quotient,* of a regular language L over an alphabet Σ by a word $w \in \Sigma^*$ is the regular language $w^{-1}L = \{x \in \Sigma^* : wx \in L\}$. It is well known that a language is regular if and only if it has a finite number of quotients. Consequently, the number of quotients of a regular language, its *quotient complexity* [1], is a natural measure of complexity of the language. Quotient complexity is also known as *state complexity* [15], which is the number of

* For a complete version of this work see http://arxiv.org/abs/1310.1841.

J. Esparza et al. (Eds.): ICALP 2014, Part II, LNCS 8573, pp. 1–12, 2014.

states in the complete minimal *deterministic finite automaton* (*DFA*) recognizing the language. We prefer quotient complexity because it is a language-theoretic concept, and we refer to it simply as *complexity*.

The problem of determining the complexity of an operation [1,8,15,16] on regular languages has received much attention. It is defined as the maximal complexity of the language resulting from the operation, taken as a function of the complexities of the operands. When operations are performed on large automata it is important to have some information about the size of the result and the time it will take to compute it. The quotient complexity of an operation gives an upper bound on its time and space complexity [15].

Languages that meet the upper bound on the complexity of an operation are *witnesses* for this operation. Although witnesses for common operations on regular languages are well known, there are occasions when one has to look for new witnesses:

1. One may be interested in a *class* of languages that have the same complexity with respect to a given operation. For example, let $\Sigma = \{a, b\}$ and let $|w|_a$ be the number of times the letter a appears in the word $w \in \Sigma^*$. Then the intersection of the languages $L = \{w \in \Sigma^* \colon |w|_a \equiv m - 1 \bmod m\}$ and $L' = \{w \in \Sigma^* \colon |w|_b \equiv n - 1 \bmod n\}$ has complexity mn. The languages $K = (b^*a)^{m-1}\Sigma^*$ and $K' = (a^*b)^{n-1}\Sigma^*$ also meet this bound; hence (L, L') and (K, K') are in the same complexity class with respect to intersection.

2. Whenever one studies complexity within a *proper subclass* of regular languages, one usually needs to find new witnesses. For example, in the class of regular right ideals – languages $L \subseteq \Sigma^*$ satisfying $L = L\Sigma^*$ – languages K and K' are appropriate, but L and L' are not. The main result of the present paper has been applied to right ideals in [4], where the proof that the witnesses used there meet the bounds for boolean operations was greatly simplified with the aid of our theorem.

3. When one studies *combined operations* – operations that involve more than one basic operation, for example, the intersection of reversed languages – one again need new witnesses [7].

Before stating our result, we provide some additional background information. The *syntactic congruence* $\leftrightarrow_L$ of L is defined as follows: For all $x, y \in \Sigma^*$, $x \leftrightarrow_L y$ if and only if $uxv \in L \Leftrightarrow uyv \in L$ for all $u, v \in \Sigma^*$. The set $\Sigma^+/\leftrightarrow_L$ of equivalence classes of the relation $\leftrightarrow_L$ is a semigroup with concatenation as the operation; it is called the *syntactic semigroup* of L, which we denote by S_L. It is well known that the syntactic semigroup is isomorphic to the semigroup $S_{\mathcal{D}}$ of transformations performed by non-empty words on the set of states in the minimal DFA $\mathcal{D}$ recognizing L; this semigroup is known as the *transition semigroup* of $\mathcal{D}$. If $\mathcal{D}$ has n states, the cardinality of the transition semigroup is bounded from above by n^n, and this bound is reachable.

The *atoms* [5,6] of a regular language are non-empty intersections of all left quotients of the language, some or all of which may be complemented. A regular language has at most 2^n atoms, and their quotient complexities are known [5].

The *reverse* of a word is defined inductively: the reverse of the empty word ε is $\varepsilon^R = \varepsilon$, and the reverse of wa with $w \in \Sigma^*$ and $a \in \Sigma$ is $(wa)^R = aw^R$. The reverse of a language L is $L^R = \{w^R \colon w \in L\}$. For L with complexity n the maximal complexity of L^R is 2^n, and this bound is reachable.

Whenever new witnesses are used, it is necessary to prove that these witnesses meet the required bound. It would be very useful to have results stating that *if the languages in question have some property P, then they meet the upper bound for a given operation.* Some results of this type are now briefly discussed.

Let **MSC** denote the class of languages with *maximal syntactic complexity* (languages with largest syntactic semigroups), let **STT** denote the class of languages whose minimal DFAs have *set-transitive transition semigroups* (for any two sets of states of the same cardinality there is a transformation that maps one set to the other), let **MAL** denote the class of *maximally atomic languages* (languages that have all 2^n atoms, all of which have maximal possible quotient complexity), let **MNA** denote the class of languages with the *maximal number* (2^n) of *atoms*, and let **MCR** denote the class of languages with a *maximally complex reverse* (reverse of complexity 2^n). The following relations hold [3]:

$$\mathbf{MSC} \subset \mathbf{STT} = \mathbf{MAL} \subset \mathbf{MNA} = \mathbf{MCR}.$$

The fact that $\mathbf{MSC} \subset \mathbf{MCR}$ is a result of A. Salomaa, Wood, and Yu [12], and the observation that $\mathbf{MNA} = \mathbf{MCR}$ was made by Brzozowski and Tamm [6].

Our main theorem relates the complexity of proper binary boolean operations on regular languages to the nature of the syntactic semigroups of the languages. A boolean operation is *proper* if it is not a constant and not a function of one variable only.

Let S_n denote the symmetric group of degree n. A *basis* [9] of S_n is an ordered pair (s, t) of distinct transformations of $Q_n = \{0, \ldots, n-1\}$ that generate S_n. Two bases (s, t) and (s', t') of S_n are *conjugate* if there exists a transformation $r \in S_n$ such that $rsr^{-1} = s'$, and $rtr^{-1} = t'$.

Assume that a DFA $\mathcal{D}$ (respectively, $\mathcal{D}'$) has state set Q_m (Q_n), and transition semigroup S_m (S_n). Let L (L') be the language accepted by $\mathcal{D}$ ($\mathcal{D}'$). Our main theorem is a generalization of a result of Brzozowski and Liu [2]:

Theorem 1. *Let $\mathcal{D}$ and $\mathcal{D}'$ be binary DFAs with m and n states respectively, where $m, n \geq 2$ and $(m, n) \notin \{(2, 2), (3, 4), (4, 3), (4, 4)\}$. If the subgroups of permutations in the transition semigroups of $\mathcal{D}$ and $\mathcal{D}'$ are S_m and S_n respectively, and $\circ$ is a proper binary boolean operation, then the complexity of $L \circ L'$ is mn, unless $m = n$ and the bases of the transition semigroups of $\mathcal{D}$ and $\mathcal{D}'$ are conjugate, in which case the quotient complexity of $L \circ L'$ is at most $m = n$.*

The proof that the complexity of a binary boolean operation is maximal involves two steps. First, one proves that the direct product of the minimal DFAs of the languages is connected, meaning that all of its states are reachable from the initial state. Second, one verifies that any two states in the direct product are distinguishable by some word, that is, that they are not equivalent. Since both reachability and distinguishability will be proved using only permutations,

it is convenient to ignore other transformations and assume that the transition semigroups of the DFAs we deal with are symmetric groups.

The remainder of the paper is structured as follows: Section 2 defines our terminology and notation. Section 3 deals with the conditions under which the direct product of two automata is connected. Section 4 studies uniformly minimal semiautomata (automata without final states), that is, semiautomata which become minimal DFAs if one adds an arbitrary set of final states, other than the empty set and the set of all states. Section 5 contains our main result relating symmetric groups to the complexity of boolean operations for all except a few cases. Section 6 concludes the paper.

2 Preliminaries

Groups. Our results rely heavily on the theory of finite groups. We refer the reader to [11,13], for example, for basic facts about groups.

Transformations. A *transformation* of a set Q is a mapping of Q into itself. We deal only with finite non-empty sets and, without loss of generality, assume that $Q = Q_n = \{0, 1, \ldots, n-1\}$. If t is a transformation of Q_n and $i \in Q_n$, then $t(i)$ is the image of i under t. An arbitrary transformation is written in the form

$$t = \begin{pmatrix} 0 & 1 & \ldots & n-2 & n-1 \\ i_0 & i_1 & \ldots & i_{n-2} & i_{n-1} \end{pmatrix},$$

where $i_k = t(k)$, $0 \leq k \leq n-1$, and $i_k \in Q_n$. The *composition* of two transformations t_1 and t_2 of Q_n is a transformation $t_1 \circ t_2$ such that $(t_1 \circ t_2)(i) = t_1(t_2(i))$ for all $i \in Q_n$. We usually omit the composition operator and write $t_1 t_2$. The set of all transformations of Q_n is a monoid under composititon with the identity transformation acting as the unit element $\mathbf{1}$.

A *permutation* is a mapping of Q_n *onto* itself. A permutation t is a *cycle of of length* k or a k*-cycle* , where $k \geq 2$, if there exist pairwise different elements $i_1, \ldots, i_k$ such that $t(i_1) = i_2$, $t(i_2) = i_3$, $\ldots$, $t(i_{k-1}) = i_k$, and $t(i_k) = i_1$, and t does not affect any other elements. A cycle is denoted by $(i_1, i_2, \ldots, i_k)$. A *transposition* is a 2-cycle. Every permutation is a product (composition) of transpositions, and the parity of the number of transpositions in the factorization is an invariant. A permutation is *odd* (*even*) if its factorization has an odd (even) number of factors. The *symmetric group* S_n of *degree* n is the set of all permutations of Q_n, with composition as the group operation, and the identity as $\mathbf{1}$. The *alternating group* A_n is the set of all even permutations of S_n.

Given a subgroup H of S_n, we say that H *acts transitively* on Q_n if for each $i, j \in Q_n$ there is some $t \in H$ such that $t(i) = j$. We say that H *acts doubly transitively* on Q_n if whenever $i, j, k, \ell \in Q_n$ with $i \neq j$ and $k \neq \ell$ there is some $t \in H$ such that $t(i) = k$, $t(j) = \ell$.

Semiautomata and Automata. A *deterministic finite semiautomaton (DFS)* is a quadruple $\mathcal{A} = (Q, \Sigma, \delta, q_0)$, where Q is a finite set of *states*, Σ is a finite non-empty *alphabet*, $\delta\colon Q \times \Sigma \to Q$ is the *transition function*, and q_0 is the *initial state*. We extend δ to $Q \times \Sigma^*$ in the usual way. A state q is *reachable* from the

initial state if there is a word w such that $q = \delta(q_0, w)$. A DFS is *connected* if every state $q \in Q$ is reachable.

For a DFS $\mathcal{A} = (Q, \Sigma, \delta, q_0)$ and a word $w \in \Sigma^*$, the transition function $\delta(\cdot, w)$ is a transformation of Q, the transformation *induced by* w. The set of all transformations induced by non-empty words is the *transition semigroup* $S_{\mathcal{A}}$ of $\mathcal{A}$. For $w \in \Sigma^+$, we denote by $w\colon t$ the transformation t of Q_n induced by w.

Given semiautomata $\mathcal{A} = (Q, \Sigma, \delta, q_0)$ and $\mathcal{A}' = (Q', \Sigma, \delta', q_0')$, we define their direct product to be the DFS $\mathcal{A} \times \mathcal{A}' = (Q \times Q', \Sigma, (\delta, \delta'), (q_0, q_0'))$.

A *deterministic finite automaton (DFA)* is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where (Q, Σ, δ, q_0) is a DFS and $F \subseteq Q$ is the set of *final states*. The DFA $\mathcal{D}$ *accepts* a word $w \in \Sigma^*$ if $\delta(q_0, w) \in F$. The set of all words accepted by $\mathcal{D}$ is the *language* $L(\mathcal{D})$ of $\mathcal{D}$. The *language accepted from a state* q of a DFA is the language $L_q(\mathcal{D})$ accepted by the DFA $(Q, \Sigma, \delta, q, F)$. Two states of a DFA are *distinguishable* if there exists a word w which is accepted from one of the states and rejected from the other. Otherwise, the two states are *equivalent*. A DFA is *minimal* if all of its states are reachable from the initial state and no two states are equivalent. Note that if $|Q| \geq 2$ and $\mathcal{D}$ is minimal, then $\emptyset \subsetneq F \subsetneq Q$.

3 Connectedness

From now on we are interested in semiautomata $\mathcal{A}$ and $\mathcal{A}'$ whose transition semigroups are symmetric groups generated by two-element bases. We assume that permutations s and s' are induced by a in $\mathcal{A}$ and $\mathcal{A}'$, and permutations t and t' by b, that is, $a\colon s$, $b\colon t$ in $\mathcal{A}$ and $a\colon s'$, $b\colon t'$ in $\mathcal{A}'$.

Example 1. Let $\Sigma = \{a, b\}$, $\mathcal{A} = (Q_3, \Sigma, \delta, 0)$, and $\mathcal{A}' = (Q_3, \Sigma, \delta', 0)$, where $a\colon s = (0, 1, 2)$, $b\colon t = (0, 1)$ in $\mathcal{A}$, and $a\colon s' = (0, 1, 2)$, $b\colon t' = (1, 2)$ in $\mathcal{A}'$. Then (s, t) and (s', t') are conjugate, since $rsr^{-1} = s'$ and $rtr^{-1} = t'$ for $r = (0, 1, 2)$. If $\mathcal{A}''$ has $s'' = (0, 1)$ and $t'' = (0, 1, 2)$, then (s, t) and (s'', t'') are not conjugate.

The transition semigroups of $\mathcal{A}$, $\mathcal{A}'$ and $\mathcal{A}''$ all have 6 elements. Those of $\mathcal{A}$ and $\mathcal{A}'$, when viewed as semigroups generated by a and b, are identical, but those of $\mathcal{A}$ and $\mathcal{A}''$ are not: for example, $a^3 = \mathbf{1}$ in $\mathcal{S}_{\mathcal{A}}$ but $a^2 = \mathbf{1}$ in $\mathcal{S}_{\mathcal{A}''}$. ■

Theorem 2. *Let $\Sigma = \{a, b\}$, let $\mathcal{A} = (Q_m, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q_n, \Sigma, \delta', 0)$ be semiautomata with transition semigroups that are symmetric groups of degrees m and n respectively, and let the corresponding bases be B and B'. For $m, n \geq 1$, the direct product $\mathcal{A} \times \mathcal{A}'$ is connected if and only if either (1) $m \neq n$ or (2) $m = n$ and B and B' are not conjugate.*

Proof. Without loss of generality, assume that $m \leq n$. Let H denote the transition semigroup of $\mathcal{A} \times \mathcal{A}'$; then H is a subgroup of $S_m \times S_n$. Define homomorphisms $\pi_1\colon H \to S_m$ and $\pi_2\colon H \to S_n$ by $\pi_1((s, t)) = s$ and $\pi_2((s, t)) = t$. Observe that π_1 and π_2 are surjective, since the transition semigroups of $\mathcal{A}$ and $\mathcal{A}'$ are S_m and S_n respectively. We let H_0 denote the subgroup of H consisting of all elements that map the set $\{0\} \times Q_n$ to itself. Then H_0 has index m in H and thus $\pi_2(H_0)$ has index at most m in $\pi_2(H) = S_n$. Thus the order of $\pi_2(H_0)$ is at least $n!/m \geq (n-1)!$.

Since a subgroup of S_n that does not act transitively on Q_n is necessarily isomorphic to a subgroup of $S_i \times S_{n-i}$ for some $i \in \{1, \ldots, n-1\}$ [14, Section 2.5.1], a subgroup of S_n whose order is strictly greater than $(n-1)!$ acts transitively on Q_n. Moreover, a subgroup of order $(n-1)!$ that does not act transitively on Q_n is isomorphic to $S_1 \times S_{n-1}$; that is, it is the stabilizer of a point. Thus $\pi_2(H_0)$ fails to act transitively on Q_n if and only if $m = n$ and $\pi_2(H_0)$ is the stabilizer of a point.

Suppose that $m < n$ or $m = n$ and $\pi_2(H_0)$ is not the stabilizer of a point, which is equivalent to assuming that $\pi_2(H_0)$ acts transitively on Q_n. We claim that the direct product $\mathcal{A} \times \mathcal{A}'$ is connected. To see this, notice that given (i, j) and (i', j') in $Q_m \times Q_n$, we can find t (respectively t') in H that sends (i, j) to $(0, k)$ (respectively (i', j') to $(0, k')$) for some k (respectively k') in Q_n, since $\pi_1(H) = S_m$ acts transitively on Q_m. Since we have assumed that $\pi_2(H_0)$ acts transitively on Q_n, we can find $t'' \in H$ such that $\pi_2(t'') \in \pi_2(H_0)$ sends $(0, k)$ to $(0, k')$. Hence $(t')^{-1}t''t$ sends (i, j) to (i', j'), and so $\mathcal{A} \times \mathcal{A}'$ is connected.

Suppose next that $m = n$ and $\pi_2(H_0)$ is the stabilizer of a point. By relabelling if necessary, we may assume that $\pi_2(H_0)$ stabilizes 0. Then H cannot send $(0, 0)$ to $(0, i)$ for $i \neq 0$ and so $\mathcal{A} \times \mathcal{A}'$ is not connected. We claim that the bases B and B' are conjugate.

To prove this claim, note that H has the property that if $(s, t) \in H \subseteq S_n \times S_n$ and $s(0) = 0$, then $t(0) = 0$. We claim there is a permutation $u \in S_n$ with $u(0) = 0$ such that if $(s, t) \in H$ sends $(0, 0)$ to (j, k), then $k = u(j)$. First suppose that $k_1, k_2 \in Q_n$ have the property that there is some $j \in Q_n$ such that (j, k_1) and (j, k_2) are in the orbit of $(0, 0)$ under the action of H. Then we can pick h in H such that $\pi_1(h)(j) = 0$. Then $(0, \pi_2(h)(k_1))$ and $(0, \pi_2(h)(k_2))$ are both in the orbit of $(0, 0)$, which means that $\pi_2(h)(k_1) = \pi_2(h)(k_2) = 0$, giving $k_1 = k_2$. It follows that there is a map $u\colon Q_n \to Q_n$ with $u(0) = 0$ such that, if $(s, t) \in H$ sends $(0, 0)$ to (j, k), then $k = u(j)$. Since $\pi_2(H)$ acts transitively on Q_n, the map u must be surjective and hence is a permutation, as claimed.

Let $s_1, s_2 \in S_n$ denote the elements in the transition semigroup corresponding to $a \in \Sigma$, and let $t_1, t_2 \in S_n$ correspond to $b \in \Sigma$. Let H' be the group generated by $(s_1, u^{-1}t_1u), (s_2, u^{-1}t_2u)$. Then H' is conjugate to H (we conjugate H by $(\mathbf{1}, u)$ to obtain H'); furthermore, H' has the property that if $(s, t) \in H'$ sends $(0, 0)$ to (i, j), then $i = j$. Thus H' acts transitively on the diagonal of $Q_n \times Q_n$; if $(s, t) \in H'$ then $s(i) = t(i)$ for all $i \in Q_n$, which gives that $s = t$. Hence, if $(s, t') \in H$, then $u^{-1}t'u = s$ and so the bases B and B' are conjugate. Thus if $\mathcal{A} \times \mathcal{A}'$ is not connected, then $m = n$ and the bases B and B' are conjugate.

Now we show the converse: If $m = n$ and the bases $B = (s, t)$ and $B' = (s', t')$ are conjugate, then $\mathcal{A} \times \mathcal{A}'$ is not connected. If $rsr^{-1} = s'$, and $rtr^{-1} = t'$, let $\psi_r\colon \{s, t\}^+ \to \{s', t'\}^+$ be the mapping that assigns to $x \in \{s, t\}^+$ the element $rxr^{-1} \in \{s', t'\}^+$. For any $x, y \in \{s, t\}^+$, if $xy = z$, then $\psi_r(x)\psi_r(y) = (rxr^{-1})(ryr^{-1}) = r(xy)r^{-1} = \psi_r(z)$. Hence the transition semigroups of $\mathcal{A}$ and $\mathcal{A}'$ are isomorphic.

The direct product $\mathcal{A} \times \mathcal{A}'$ is defined by $(Q_n \times Q_n, \{a, b\}, (\delta, \delta'), (0, 0))$, where $(\delta, \delta')((i, j), a) = (s(i), rsr^{-1}(j))$ and $(\delta, \delta')((i, j), b) = (t(i), rtr^{-1}(j))$ for any $i, j \in Q_n$.

If $\mathcal{A} \times \mathcal{A}'$ is connected, then for all $(i, j) \in Q_n \times Q_n$ there must exist a word $w \in \Sigma^+$ such that $(\delta, \delta')((0, 0), w) = (i, j)$ or, equivalently, there exists a permutation p such that $p(0) = i$ and $rpr^{-1}(0) = j$. There are now two cases:

1. If $r^{-1}(0) \neq 0$, we prove that state $(i, r(i))$ is unreachable for all $i \in Q_n$. If $(i, r(i))$ is reachable, then there exists a permutation p such that $p(0) = i$ and $rpr^{-1}(0) = r(i)$. But then $r^{-1}rpr^{-1}(0) = pr^{-1}(0) = i = p(0)$, and so $p^{-1}pr^{-1}(0) = r^{-1}(0) = 0$, which is a contradiction.

2. If $r^{-1}(0) = 0$, we prove that state (i, i) is unreachable for some $i \in Q_n$. Since r cannot be the identity, there must exist an i such that $r(i) \neq i$. Suppose (i, i) is reachable for that i. Then there exists a permutation p such that $p(0) = i$ and $rpr^{-1}(0) = i$. Thus $i = rpr^{-1}(0) = rp(0) = p(0)$ and $r(i) = i$, which is a contradiction.

Hence $\mathcal{A} \times \mathcal{A}'$ cannot be connected. □

Remark 1. If $\mathcal{A} \times \mathcal{A}'$ is connected, then it is strongly connected, since the transition semigroup of $\mathcal{A} \times \mathcal{A}'$ is a group.

4 Uniformly Minimal Semiautomata

Semiautomata that result in minimal DFAs under any non-trivial assignment of final states were studied by Restivo and Vaglica [10]. We modify their definitions slightly to suit our purposes. A strongly connected DFS $\mathcal{A} = (Q, \Sigma, \delta, q_0)$ with $|Q| \geq 2$ is *uniformly minimal* if the DFA $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$ is minimal for each set F of final states, where $\emptyset \subsetneq F \subsetneq Q$.

Given a DFS $\mathcal{A} = (Q, \Sigma, \delta, q_0)$, we define the *pair graph* of $\mathcal{A}$ to be the directed graph $G_{\mathcal{A}} = (V_{\mathcal{A}}, E_{\mathcal{A}})$, where the set $V_{\mathcal{A}}$ of vertices is the set of all two-element subsets $\{p, q\}$ of Q, and the set $E_{\mathcal{A}}$ of edges consists of unordered pairs $(\{p, q\}, \{p', q'\})$ such that $\{\delta(p, a), \delta(q, a)\} = \{p', q'\}$. The following is from [10]:

Proposition 1 (Restivo and Vaglica). *Let $\mathcal{A} = (Q, \Sigma, \delta, q_0)$ be a strongly connected DFS with at least two states. If the pair graph $(V_{\mathcal{D}}, E_{\mathcal{D}})$ is strongly connected, then $\mathcal{A}$ is uniformly minimal.*

We prove a similar result for semiautomata with symmetric groups.

Proposition 2. *Suppose that $\mathcal{A} = (Q_n, \Sigma, \delta, q_0)$ is a DFS and the transition semigroup $S_{\mathcal{A}}$ of $\mathcal{A}$ is the symmetric group S_n. Then $\mathcal{A}$ is strongly connected and uniformly minimal.*

Proof. If $S_{\mathcal{A}} = S_n$, then $S_{\mathcal{A}}$ contains all permutations of Q_n, in particular, the cycle $(0, \ldots, n-1)$; hence $\mathcal{A}$ is strongly connected. For any $(i, j), (k, \ell) \in Q_n \times Q_n$ with $i \neq j$, $k \neq \ell$, and $\{i, j\} \neq \{k, \ell\}$, any permutation that maps i to k and j to ℓ connects $\{i, j\}$ to $\{k, \ell\}$ in the pair graph of $\mathcal{A}$. Hence the pair graph is strongly connected, and $\mathcal{A}$ is uniformly minimal by Proposition 1. □

Let the truth values of propositions be 1 (true) and 0 (false). Let $\circ\colon \{0,1\} \times \{0,1\} \to \{0,1\}$ be a binary boolean function. Extend $\circ$ to a function $\circ\colon 2^{\Sigma^*} \times 2^{\Sigma^*} \to 2^{\Sigma^*}$: If $w \in \Sigma^*$ and $L, L' \subseteq \Sigma^*$, then $w \in (L \circ L') \Leftrightarrow (w \in L) \circ (w \in L')$. Also, extend $\circ$ to a function $\circ\colon 2^{Q_m} \times 2^{Q_n} \to 2^{Q_m \times Q_n}$: If $q \in Q_m$, $q' \in Q_n$, $F \subseteq Q_m$, and $F' \subseteq Q_n$, then $(q, q') \in (F \circ F') \Leftrightarrow (q \in F) \circ (q' \in F')$.

Suppose that $\mathcal{A} = (Q, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q', \Sigma, \delta', 0)$ with $|Q| = m$ and $|Q'| = n$ are uniformly minimal DFSs, and $\circ$ is any proper boolean function. The pair $(\mathcal{A}, \mathcal{A}')$ is *uniformly minimal for* $\circ$ if the direct product $\mathcal{P} = (Q \times Q', \Sigma, (\delta, \delta'), (0,0), F \circ F')$ is minimal for all *valid assignments* of sets F and F' of final states to $\mathcal{A}$ and $\mathcal{A}'$, that is, sets such that $\emptyset \subsetneq F \subsetneq Q$ and $\emptyset \subsetneq F' \subsetneq Q'$.

If $n = 1$, then $\mathcal{A} \times \mathcal{A}'$ is isomorphic to $\mathcal{A}$ and no boolean function $\circ$ is proper. Hence this case, and also the case with $m = 1$, is of no interest. Henceforth we assume that $m, n \geq 2$.

We now consider pair graphs of DFSs with symmetric groups as their transition semigroups.

Example 2. Suppose that $m = n = 2$, and $\mathcal{A}$ and $\mathcal{A}'$ both have S_2 as their transition semigroup. There are two permutations in S_2: $(0,1)$ and $\mathbf{1}$, and there are three bases: $B_1 = (a\colon (0,1), b\colon (0,1))$, $B_2 = (a\colon (0,1), b\colon \mathbf{1})$, and $B_3 = (a\colon \mathbf{1}, b\colon (0,1))$. Note that no two of these bases are conjugate.

For each basis, there are two possible final states, 0 or 1, and hence two DFAs; thus there are six different DFAs. There are then twelve direct products $\mathcal{D}_j^i \times \mathcal{D}_\ell^k$ with non-conjugate bases, where $\mathcal{D}_j^i$ ($\mathcal{D}_\ell^k$) uses basis B_i (B_k) and has j (ℓ) as final state, for $i, k = 1, 2, 3$ and $j, \ell = 1, 2$.

For each pair of DFAs accepting languages L and L' respectively, we tested the complexity of five boolean functions: $L \cup L'$, $L \cap L'$, $L \oplus L'$, $L \setminus L'$ and $L' \setminus L$. Note that the complexity of each remaining proper boolean function is the same as that of one of these five functions. For all twelve direct products of DFAs with non-conjugate bases, all proper boolean functions reach the maximal complexity 4, except for the functions $L \oplus L'$ and $\overline{L \oplus L'}$, which fail in all twelve cases. Thus any two DFAs $\mathcal{D} = (Q_2, \Sigma, \delta_i, 0, F)$ and $\mathcal{D}' = (Q_2, \Sigma, \delta_k, 0, F')$, where $Q_2 = \{0,1\}$, $\Sigma = \{a, b\}$, δ_i (δ_k) is defined by basis B_i (B_k), $F = \{j\}$ and $F' = \{\ell\}$, are uniformly minimal for all proper boolean functions, except $\oplus$ and its complement. So our main result applies only in some cases if $m = n = 2$. ■

Proposition 3. *Let $\mathcal{A} = (Q_m, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q_n, \Sigma, \delta', 0)$, with $m, n \geq 2$ and $\max(m, n) \geq 3$, be DFSs with transition semigroups that are symmetric groups, and let $\mathcal{P}$ be their direct product. Then the following hold:*

1. The pair graph of $\mathcal{P}$ consists of strongly connected components – which we will call simply components *– of one of the following three types:*

(a) $T_1 \subseteq C_1 = \{\{(i,j),(k,\ell)\}\colon i \neq k, j \neq \ell\}$,

(b) $T_2 \subseteq C_2 = \{\{(i,j),(i,\ell)\}\colon j \neq \ell\}$,

(c) $T_3 \subseteq C_3 = \{\{(i,j),(k,j)\}\colon i \neq k\}$.

2. Every state (i,j) of the direct product $\mathcal{P}$ appears in at least one pair in each component.

3. Each component has at least $mn/2 \geq 3$ pairs.

Proof. The first claim follows since the transition semigroup of $\mathcal{P}$ is a group. The second claim holds because the direct product is strongly connected, by Remark 1. For the third claim, note that there are mn states in $\mathcal{P}$, but they can appear in pairs; hence the bound $mn/2$. Since we are assuming that $mn \geq 6$, the last claim follows. □

Now consider DFAs $\mathcal{D} = (Q_m, \Sigma, \delta, 0, F)$ and $\mathcal{D}' = (Q_n, \Sigma, \delta', 0, F')$, where $\emptyset \subsetneq F \subsetneq Q_m$ and $\emptyset \subsetneq F' \subsetneq Q_n$. A state $\{(i,j),(k,\ell)\}$ of the pair graph of the direct product $\mathcal{P}$ of $\mathcal{D}$ and $\mathcal{D}'$ is *distinguishing* if and only if (i,j) is final and (k,ℓ) is not, or *vice versa*.

Example 3. Suppose $m = 3$, $n = 4$, δ is defined by the basis $(a\colon (0,1), b\colon (0,1,2))$ of S_3, and δ' by the basis $(a\colon (0,1), b\colon (1,3,2))$ of S_4. One verifies that these bases are not conjugate. The direct product $\mathcal{P}$ is connected and has twelve states.

If $F = \{2\}$, $F' = \{0,1\}$ and intersection is the boolean function, then there are no distinguishing pairs in the component of the pair graph T containing $\{(0,0),(0,3)\}$. Hence any two states appearing in the same pair of T are equivalent. Indeed, the minimal version of $\mathcal{P}$ has only six states. ■

Example 4. Suppose $m = n = 4$, δ is defined by the basis $(a\colon (0,1,2), b\colon (2,3))$, and δ' by the basis $(a\colon (1,3,2), b\colon (0,2,1,3))$. If $F = \{0,1\}$ and $F' = \{0,1\}$, then the complexity of $L \oplus L'$ is 4, but all the other complexities are 12. ■

Lemma 1. *Let $\mathcal{D} = (Q, \Sigma, \delta, 0, F)$ and $\mathcal{D}' = (Q', \Sigma, \delta', 0, F')$, with $|Q|, |Q'| \geq 2$, be DFAs with transition semigroups that are groups, and let $\mathcal{P} = (Q \times Q', \Sigma, (\delta, \delta'), (0,0), F \circ F')$ be their direct product. Then $\mathcal{P}$ is minimal if and only if every component of the pair graph $G_{\mathcal{P}}$ of $\mathcal{P}$ has a distinguishing pair.*

5 Symmetric Groups and Boolean Operations

We begin with a well-known but apparently unpublished result.

Lemma 2. *Let n be a positive integer, let G be either S_n or A_n, and let H be a subgroup of G of index $m \leq n$. Then the following hold:*

(i) if $n \neq 4$ and $m < n$, then H is either A_n or S_n;

(ii) if $m = n$ and $n \neq 6$, then there is some $i \in Q_n$ such that H is the set of permutations in G that fix i.

(iii) if $m = n = 6$, then there is an automorphism ϕ of S_6 such that $\phi(H)$ is the set of elements that fix 0.

The following lemma, like Theorem 2, deals with reachability. The conditions in the lemma, however, are useful for determining reachability in the pair graph of $\mathcal{A} \times \mathcal{A}'$, rather than in $\mathcal{A} \times \mathcal{A}'$ itself.

Lemma 3. *Let $\Sigma = \{a,b\}$, let $\mathcal{A} = (Q_m, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q_n, \Sigma, \delta', 0)$ be semiautomata with transition semigroups that are symmetric groups of degrees m and n respectively with $m \leq n$, $n \neq 4$ and $(m,n) \neq (6,6)$. Let H be the transition semigroup of $\mathcal{A} \times \mathcal{A}'$, and let π_1 and π_2 be the natural projections from H onto S_m and S_n respectively. If $H_0 = \{h \in H\colon \pi_1(h)(0) = 0\}$, then*

1. $\pi_2(H_0)$ is either S_n or A_n, or is the stabilizer of a point in Q_n.

2. $\pi_2(H_0)$ is the stabilizer of a point if and only if $m = n$, and in this case the direct product $\mathcal{A} \times \mathcal{A}'$ is not connected.

Proof. For Part 1, since $\pi_1(H) = S_m$, for each $i \in \{0, \ldots, m-1\}$ there is some $h_i \in H$ such that $\pi_1(h_i)$ takes 0 to i. For a given $h \in H$, $\pi_1(h)$ takes 0 to j for some $j \in \{0, 1, \ldots, m-1\}$, and thus $h_j^{-1}h \in H_0$ and so $h \in h_j H_0$. However, since $\pi_1(h)$ takes 0 to j, we have $h_i^{-1}h \notin H_0$ and thus $h \notin h_i H_0$ for $i \neq j$. Thus the cosets $h_0H, \ldots, h_{m-1}H$ are distinct, and H_0 has index m in H. Since $\pi_2(H) \subseteq \bigcup_{i=0}^{m-1} \pi_2(h_i)\pi_2(H_0)$, $\pi_2(H_0)$ has index at most m in $\pi_2(H) = S_n$. If $n \neq 4$ and $m < n$, then $\pi_2(H_0)$ is either A_n or S_n by Lemma 2. If $m = n$ and $n \neq 6$, then $\pi_2(H_0)$ has index n in S_n and hence must be the stabilizer of a some $i \in Q_n$ by Lemma 2.

For Part 2, suppose that $m = n$ and $\pi_2(H_0)$ is the stabilizer of a point in Q_n. By relabelling if necessary, we may assume that $\pi_2(H_0)$ stabilizes 0. Hence, if $h \in H$ sends $(0,0)$ to $(0,j)$ then $j = 0$. In particular, there is no $h \in H$ that sends $(0,0)$ to $(0,1)$ or that sends $(0,1)$ to $(0,0)$, and so $\mathcal{A} \times \mathcal{A}'$ is necessarily not connected. □

Lemma 4. *Let $\mathcal{A} = (Q_m, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q_n, \Sigma, \delta', 0)$ be semiautomata with transition semigroups that are the symmetric groups of degrees m and n, respectively with $m \leq n$, $m \geq 2$, $n \geq 5$, and $(m,n) \neq (6,6)$. If $\mathcal{A} \times \mathcal{A}'$ is connected, then the pair graph of $\mathcal{A} \times \mathcal{A}'$ has exactly three connected components: $C_1 = \{\{(i,j),(k,\ell)\} \colon i \neq k, j \neq \ell\}$, $C_2 = \{\{(i,j),(i,\ell)\} \colon j \neq \ell\}$, and $C_3 = \{\{(i,j),(k,j)\} \colon i \neq k\}$.*

Proof. We let H denote the transition semigroup of $\mathcal{A} \times \mathcal{A}'$. We show that each of C_1, C_2, C_3 is strongly connected. Note that each of C_1, C_2, C_3 is necessarily a union of connected components.

We show that C_1 is strongly connected. Suppose we have pairs $\{(i,j),(k,\ell)\}$ and $\{(i',j'),(k',\ell')\}$ with i,k distinct, i',k' distinct, j,ℓ distinct, and j',ℓ' distinct. Since S_m acts doubly transitively on Q_m when $m \geq 2$, there is some $s \in H$ that sends (i,j) to (i',j'') and (k,ℓ) to (k',ℓ'') for some $j'',\ell'' \in Q_n$.

Thus we may assume without loss of generality that $i' = i$ and $k' = k$. Let H_0 be the subgroup of $S_m \times S_n$ consisting of all $x \in H$ such that $\pi_1(x)$ fixes i. By Lemma 3, since we assume that $\mathcal{A} \times \mathcal{A}'$ is connected, $\pi_2(H_0)$ is not a stabilizer of a point in Q_n. Hence $\pi_2(H_0)$ is either S_n or A_n. Let H_1 denote the subgroup of $S_m \times S_n$ consisting of all $x \in H$ such that $\pi_1(x)$ fixes i and k. By the argument used in Lemma 3 to show that $\{h \in H \colon \pi_1(h)(0) = 0\}$ has index m in H, we see that $\pi_2(H_1)$ has index at most $m-1$ in $\pi_2(H_0)$. Thus $\pi_2(H_1)$ is a subgroup of A_n or S_n of index at most $n-1$, and hence must again be A_n or S_n by Lemma 2. Since A_n and S_n both act doubly transitively on Q_n, there is some $h \in H$ that sends (i,j) to (i,j') and (k,ℓ) to (k,ℓ') whenever ℓ and ℓ' are distinct. This proves that C_1 is indeed a strongly connected component.

Next, consider pairs $\{(i,j),(i,k)\}$ with j,k distinct. For given $\{(i',j'),(i',k')\}$ with j',k' distinct, there is some element $s \in H$ such that $\pi_1(s)(i) = i'$ and thus s sends (i,j) to (i',j'') and (i,k) to (i',k'') for some $j'',k'' \in Q_n$ with $j'' \neq k''$.

Now note that $\pi_2(\{x \in H\colon \pi_1(x)(i') = i'\})$ is either S_n or A_n by Lemma 3, and thus acts doubly transitively on Q_n. It follows that there is some $s' \in H$ such that s' sends (i', j'') to (i', j') and (i', k'') to (i', k'). Then $s's$ sends $\{(i,j),(i,k)\}$ to $\{(i',j'),(i',k')\}$ and thus C_2 is strongly connected.

Finally, consider pairs $\{(i,j),(k,j)\}$ and $\{(i',j'),(k',j')\}$ with i,k distinct and i',k' distinct. From the argument used in proving C_1 is strongly connected, we see that we can find $s \in H$ that sends $\{(i,j),(k,j)\}$ to $\{(i',j''),(k',j'')\}$ for some j''. As in the proof that C_1 is strongly connected, we see that the image of the set of $h \in H$ for which $\pi_1(h)$ stabilizes both i' and k' under π_2 acts transitively on Q_n; hence we can find $s' \in H$ that sends $\{(i',j''),(k',j'')\}$ to $\{(i',j'),(k',j')\}$. Thus C_3 is strongly connected. □

Corollary 1. *Let m and n be positive integers with $n \geq m \geq 2$, $n \geq 5$, and $(m,n) \neq (6,6)$, and let $\mathcal{A} = (Q_m, \Sigma, \delta, 0)$ and $\mathcal{A}' = (Q_n, \Sigma, \delta', 0)$ be semiautomata with transition semigroups that are the symmetric groups of degrees m and n. Suppose that the direct product $\mathcal{A} \times \mathcal{A}'$ is connected and assume further that sets of final states are added to $\mathcal{A}$ and $\mathcal{A}'$ and that $\circ$ is a proper binary boolean function that defines the set of final states of the direct product $\mathcal{P}$. Then $\mathcal{P}$ is minimal for any such $\circ$.*

Proof. By Lemma 4, the pair graph of $\mathcal{A} \times \mathcal{A}'$ has three strongly connected components: $C_1 = \{\{(i,j),(k,\ell)\}\colon i \neq k, j \neq \ell\}$, $C_2 = \{\{(i,j),(i,\ell)\}\colon j \neq \ell\}$, and $C_3 = \{\{(i,j),(k,j)\}\colon i \neq k\}$.

For $(i,j) \in Q_m \times Q_n$, define $f((i,j))$ to be 1 if (i,j) is a final state, and 0, otherwise. We first claim that C_1 has a distinguishing pair, that is, there are pairs (i,j) and (k,ℓ) in $Q_m \times Q_n$ with $i \neq k$ and $j \neq \ell$ such that $f((i,j)) \neq f((k,\ell))$.

Suppose no distinguishing pair exists in C_1. Assume without loss of generality that $f((0,0)) = 0$. then $f((i,j)) = 0$ whenever $i \neq 0$ and $j \neq 0$. Given $k \in Q_n$, we pick $\ell \in Q_n \setminus \{0,k\}$; this is always possible since $n \geq 3$. Since $\{(0,k),(1,\ell)\}$ is in C_1 and we have assumed that C_1 has no distinguishing pairs, we must have $f((0,k)) = f((1,\ell))$. But $f(1,\ell)$ must be 0, for otherwise we would have the distinguishing pair $\{(0,0),(1,\ell)\}$. Hence $f((0,k)) = f((1,\ell)) = 0$. Thus we have $f((i,j)) = 0$ for every $i \in Q_m$ and every $j \in Q_n \setminus \{0\}$. Similarly, we must have $f((i,0)) = f((0,1)) = 0$ for $i \in Q_m \setminus \{0\}$, and hence f is the zero function, a contradiction.

The fact that C_2 and C_3 both have distinguishing pairs follows from the fact that $\circ$ is a proper boolean function. By Lemma 1, we conclude that $\mathcal{A} \times \mathcal{A}'$ is uniformly minimal. □

We have proved our main result in the case that $m \leq n$ and $n \geq 5$ if $(m,n) \neq (6,6)$. By symmetry we may always assume that $m \leq n$. The case $(m,n) = (2,2)$ was handled in Example 2, that of $(m,n) = (3,4)$, in Example 3, and that of $(m,n) = (4,4)$, in Example 4. So the only cases to consider are those with $(m,n) \in \{(2,3),(2,4),(3,3),(6,6)\}$; these cases are covered at `http://arxiv.org/abs/1310.1841`.

6 Conclusions

We have shown that if the inputs of two DFAs induce transformations that constitute non-conjugate bases of symmetric groups, then the quotient complexity of all non-trivial boolean operations on the languages accepted by the DFAs is maximal, except for a few special cases when the sizes of the DFAs are small. We believe that other similar results are possible and deserve further study.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada under grants No. 611456 and OGP0000871, by the European Regional Development Fund through the programme COMPETE, and by the Portuguese Government through the FCT under projects PEst-C/MAT/UI0144/2011 and CANTE-PTDC/EIA-CCO/101904/2008. We thank Gareth Davies for his careful proofreading and constructive comments.

References

1. Brzozowski, J.: Quotient complexity of regular languages. J. Autom. Lang. Comb. 15(1/2), 71–89 (2010)
2. Brzozowski, J.: In search of the most complex regular languages. Int. J. Found. Comput. Sc. 24(6), 691–708 (2013)
3. Brzozowski, J., Davies, G.: Maximally atomic languages. In: Ésik, Z., Fülop, Z. (eds.) 14th International Conference Automata and Formal Languages, AFL 2014, Szeged, Hungary, May 27-29. EPTCS, vol. 151, pp. 151–161 (2014)
4. Brzozowski, J., Davies, G.: Most complex regular right-ideal languages. In: 16th International Workshop on Descriptional Complexity of Formal Systems, DCFS 2014, Turku, Finland, August 5-8. LNCS 8614 (to appear, 2014)
5. Brzozowski, J., Tamm, H.: Complexity of atoms of regular languages. Int. J. Found. Comput. Sc. 24(7), 1009–1027 (2013)
6. Brzozowski, J., Tamm, H.: Theory of átomata. Theoret. Comput. Sci. (article in press, 2014)
7. Liu, G., Martin-Vide, C., Salomaa, A., Yu, S.: State complexity of basic language operations combined with reversal. Inform. and Comput. 206, 1178–1186 (2008)
8. Maslov, A.N.: Estimates of the number of states of finite automata. Dokl. Akad. Nauk SSSR 194, 1266–1268 (1970) (Russian); English Translation: Soviet Math. Dokl. 11, 1373–1375 (1970)
9. Piccard, S.: Sur les bases du groupe symétrique. Časopis Pro Pěstování Matematiky a Fysiky 68(1), 15–30 (1939)
10. Restivo, A., Vaglica, R.: A graph theoretic approach to automata minimality. Theoret. Comput. Sc. 429, 282–291 (2012)
11. Rotman, J.: The Theory of Groups: An Introduction. Allyn and Bacon, Inc., Boston (1965)
12. Salomaa, A., Wood, D., Yu, S.: On the state complexity of reversals of regular languages. Theoret. Comput. Sci. 320, 315–329 (2004)
13. Suzuki, M.: Group Theory, vol. 1. Springer, Berlin (1982)
14. Wilson, R.: The Finite Simple Groups. Springer, Berlin (2009)
15. Yu, S.: State complexity of regular languages. J. Autom. Lang. Comb. 6, 221–234 (2001)
16. Yu, S., Zhuang, Q., Salomaa, K.: The state complexities of some basic operations on regular languages. Theoret. Comput. Sci. 125(2), 315–328 (1994)