AUTOMORPHISM GROUPS OF SQUARES AND OF FREE ALGEBRAS

KEITH A. KEARNES AND STEVEN T. TSCHANTZ

Abstract. We show that certain finite groups do not arise as the automorphism group of the square of a finite algebraic structure, nor as the automorphism group of a finite, 2-generated, free, algebraic structure.

1. Introduction

Every group can be represented as the automorphism group of some algebra. For example, the regular G-set $\langle G; G\rangle$ has automorphism group isomorphic to G. In [4], Matthew Gould determined which groups are representable as the automorphism group of the square of some algebra. $\mathrm{Aut}(\mathbf{A}^2)$ contains the map $\sigma\colon (x, y) \mapsto (y, x)$ that switches coordinates, which is an involution. Gould showed that, conversely, if G is any group with an involution σ, then there is an algebra $\mathbf{A}$ such that $\mathrm{Aut}(\mathbf{A}^2) \cong G$ where σ corresponds to the map that switches the coordinates. Gould's construction produces an infinite algebra even when G is finite, which led him to consider the question of which groups are representable as the automorphism group of the square of a finite algebra. A sufficient condition for representability was supplied by Gould in [5]: If G is a finite group with involution σ and G has a retraction onto the subgroup $\{1, \sigma\}$, then there is a finite algebra $\mathbf{A}$ where $\mathrm{Aut}(\mathbf{A}^2) \cong G$ and σ represents the automorphism that switches the coordinates. Gould showed that this sufficient condition is not a necessary one.

There is an equivalent version of this representability problem, which is the version that will be considered in this paper. An algebra will be called *free* if it is a free algebra in the variety that it generates. The automorphism group of a 2-generated free algebra also has a distinguished involution, namely the automorphism induced by switching the two free generators. Gould showed in [5] that a group G with involution σ is representable as $\mathrm{Aut}(\mathbf{F})$ for some 2-generated free algebra $\mathbf{F}$, with σ representing the automorphism that switches the generators, if and only if $G \cong \mathrm{Aut}(\mathbf{A}^2)$ for some algebra $\mathbf{A}$, with σ representing the automorphism that switches the coordinates. He showed, moreover, that $\mathbf{F}$ can be taken to be finite if and only if $\mathbf{A}$ can be taken to

1991 *Mathematics Subject Classification.* Primary: 20B25; Secondary: 08A35, 08B20.
Key words and phrases. Finite automorphism groups, free algebras.

be finite. Thus the existence of a finite algebra $\mathbf{A}$ with $G \cong \mathrm{Aut}(\mathbf{A}^2)$ is equivalent to the existence of a finite, 2-generated, free algebra $\mathbf{F}$ with $G \cong \mathrm{Aut}(\mathbf{F})$.

The smallest group whose representability as $\mathrm{Aut}(\mathbf{F})$ for finite $\mathbf{F}$ is not decided by the results of Gould is therefore $\mathbb{Z}_4$. The problem of deciding the representability of this group was widely circulated for many years as the "$\mathbb{Z}_4$ Problem". The surprising fact that $\mathbb{Z}_4$ is not representable as $\mathrm{Aut}(\mathbf{F})$ when $\mathbf{F}$ is finite was announced by the second author in 1996, but not published. Since then, $\mathbb{Z}_4$ has been the only known nonrepresentable group of even order. The purpose of this paper is to identify several infinite families of nonrepresentable groups of even order. (This paper also serves to record the solution of the $\mathbb{Z}_4$ Problem.)

The groups that we consider will typically have a designated involution, always denoted by σ, which in the automorphism group of a free algebra is the automorphism that switches the two free generators. A group homomorphism φ that is assumed to satisfy $\varphi(\sigma) = \sigma$ will be written $\varphi\colon G_\sigma \to H_\sigma$, while a homomorphism not assumed to preserve σ will be written $\varphi\colon G \to H$. Similarly, $G_\sigma \cong H_\sigma$ means that there is a pointed group isomorphism from G to H. Whether or not G_σ is representable as $\mathrm{Aut}(\mathbf{F})_\sigma$ depends only on the isomorphism type of G_σ, in particular only on the conjugacy class of σ. Throughout the paper, automorphisms act on the right, while functions, operations and even other homomorphisms act on the left.

We approach the problem of determining which groups G_σ arise as $\mathrm{Aut}(\mathbf{F})_\sigma$ for some finite, 2-generated, free algebra $\mathbf{F}$ in the following way. First, we argue that if G_σ is representable as $\mathrm{Aut}(\mathbf{F})_\sigma$, then $\mathbf{F}$ may be taken to belong to a certain variety $\mathcal{V}[G_\sigma]$, defined in the next section. Then we consider a minimal subvariety $\mathcal{M}$ of the variety generated by $\mathbf{F}$. If $\mathbf{E}$ is the 2-generated free algebra in $\mathcal{M}$, then there is a natural surjective homomorphism $\nu\colon \mathbf{F} \to \mathbf{E}$. This algebra homomorphism induces a pointed group homomorphism $\widehat{\nu}\colon \mathrm{Aut}(\mathbf{F})_\sigma \to \mathrm{Aut}(\mathbf{E})_\sigma$. By partially classifying the possibilities for $\mathcal{M}$, $\mathbf{E}$, and $\mathrm{im}(\widehat{\nu})$, we are able to show that certain groups G_σ cannot arise as $\mathrm{Aut}(\mathbf{F})_\sigma$. This approach allows us to completely settle the question of which pointed groups G_σ with σ in the center of G are representable as $\mathrm{Aut}(\mathbf{F})_\sigma$ when $\mathbf{F}$ is finite (Corollary 6.4). When σ does not lie in the center of G, we obtain an assortment of nonrepresentability results involving alternating groups, special linear groups, Suzuki groups, Mathieu groups, and groups whose 2-Sylow subgroups are generalized quaternion.

2. Congruence Permutability

A variety $\mathcal{V}$ of algebras is *congruence permutable* if whenever α and β are congruences on some algebra $\mathbf{A} \in \mathcal{V}$, then $\alpha \circ \beta = \beta \circ \alpha$. This property of $\mathcal{V}$ is equivalent to the existence of a ternary term p of $\mathcal{V}$ such that the equations $p(x, y, y) \approx x$ and $p(x, x, y) \approx y$ hold in $\mathcal{V}$. Such a p is called a *Maltsev term* for $\mathcal{V}$. In this section we

establish the fact that certain pointed groups G_σ are representable as $\mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$ only when $\mathcal{V}$ is congruence permutable.

Definition 2.1. Let G be a group with a designated involution σ. $\mathcal{V}[G_\sigma]$ is the variety with binary operation symbols $\{t_\alpha \mid \alpha \in G\}$ defined by the equations

(i) $t_1(x,y) \approx x$,
(ii) $t_\sigma(x,y) \approx y$, and
(iii) $t_{\alpha\beta}(x,y) \approx t_\alpha(t_\beta(x,y), t_{\sigma\beta}(x,y))$.

These equations entail $t_{\alpha\sigma}(x,y) \approx t_\alpha(y,x)$.

Lemma 2.2. *Let $\mathbf{F}$ be free on two generators in some subvariety $\mathcal{V} \leq \mathcal{V}[G_\sigma]$. The endomorphism ε_α of $\mathbf{F}$ that is defined on the generators by $x \mapsto t_\alpha(x,y)$ and $y \mapsto t_{\sigma\alpha}(x,y)$ is an automorphism of $\mathbf{F}$, and $\kappa\colon G_\sigma \to \mathit{Aut}(\mathbf{F})_\sigma\colon \alpha \mapsto \varepsilon_\alpha$ is a homomorphism of pointed groups.*

Proof. Equations (i) and (ii) of Definition 2.1 guarantee that ε_1 is the identity on the generators of $\mathbf{F}$ and that ε_σ is the endomorphism defined on generators by $x \mapsto y$ and $y \mapsto x$ (or, more simply, $(x,y) \mapsto (y,x)$). Equation (iii) of Definition 2.1 guarantees that $\varepsilon_{\alpha\beta}$ agrees with $\varepsilon_\alpha \circ \varepsilon_\beta$ on the generators. Therefore κ is a σ-preserving monoid homomorphism from G_σ to $\mathrm{End}(\mathbf{F})$. This forces each ε_α to be invertible, with inverse $\varepsilon_{\alpha^{-1}}$, and also forces κ to be a pointed group homomorphism from G_σ to $\mathrm{Aut}(\mathbf{F})_\sigma$. □

We call the homomorphism κ of Lemma 2.2 the *canonical* homomorphism from G_σ to $\mathrm{Aut}(\mathbf{F})_\sigma$.

Theorem 2.3. *Let $\mathcal{W}$ be a variety, $\mathbf{F} = \mathbf{F}_\mathcal{W}(x,y)$, and $G_\sigma = \mathit{Aut}(\mathbf{F})_\sigma$. There is a variety $\mathcal{V}$ that is a reduct of $\mathcal{W}$ and a subvariety $\mathcal{V}[G_\sigma]$ such that the canonical homomorphism $\kappa\colon G_\sigma \to \mathit{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$ is an isomorphism.*

Proof. For each $\alpha \in G_\sigma = \mathrm{Aut}(\mathbf{F})_\sigma$, let t_α be a binary $\mathcal{W}$-term for which $t_\alpha(x,y) := x\alpha \in F$ is the image of the generator $x \in F$ under the automorphism α. Clearly $t_1(x,y) = x$ and $t_\sigma(x,y) = y$ in $\mathbf{F}$. Moreover, since $y\alpha = (x\sigma)\alpha = x(\sigma\alpha) = t_{\sigma\alpha}(x,y)$, the generators x and y are mapped by α to $t_\alpha(x,y)$ and $t_{\sigma\alpha}(x,y)$ respectively. From this it follows that

$$t_{\alpha\beta}(x,y) = x\alpha\beta = t_\alpha(x,y)\beta = t_\alpha(x\beta, y\beta) = t_\alpha(t_\beta(x,y), t_{\sigma\beta}(x,y)).$$

The equalities that we have established among the terms $\{t_\alpha \mid \alpha \in \mathrm{Aut}(\mathbf{F})\}$ imply that equations (i)—(iii) of Definition 2.1 hold in $\mathcal{W}$.

Let $\mathbf{F}'$ be the reduct of $\mathbf{F}$ to the terms $\{t_\alpha \mid \alpha \in \mathrm{Aut}(\mathbf{F})\}$ and let $\mathcal{V} = \mathsf{HSP}(\mathbf{F}')$. $\mathcal{V}$ is a reduct of $\mathcal{W}$ because $\mathbf{F}'$ is a reduct of $\mathbf{F}$. The equations established in the previous paragraph hold in $\mathbf{F}'$, hence in $\mathcal{V}$, so $\mathcal{V} \leq \mathcal{V}[G_\sigma]$.

We claim that the subalgebra $\mathbf{F}'' \leq \mathbf{F}'$ generated by $\{x,y\}$ is free over $\{x,y\}$ in $\mathcal{V}$. Since $\mathcal{V}$ is generated by $\mathbf{F}'$, $\mathbf{F}''$ is a subalgebra of $\mathbf{F}'$, and $\mathbf{F}''$ is generated by $\{x,y\}$, to prove this we must show that every function $f\colon \{x,y\} \to F'$ extends to

a homomorphism from $\mathbf{F}''$ to $\mathbf{F}'$. If $f\colon \{x,y\} \to F' = F$ is a function, then the freeness of $\mathbf{F}$ implies that f extends to an endomorphism $\widehat{f}\colon \mathbf{F} \to \mathbf{F}$. Thus $\widehat{f}|_{F''}$ is an extension of f to a homomorphism of $\mathbf{F}''$ to $\mathbf{F}'$.

To see why the canonical homomorphism $\kappa\colon G_\sigma \to \mathrm{Aut}(\mathbf{F}'')_\sigma$ is injective, assume that $\alpha \in \ker(\kappa)$. Then $\kappa(\alpha) = \varepsilon_\alpha$ is the identity automorphism of $\mathbf{F}''$. Since ε_α is defined on generators by $(x,y) \mapsto (t_\alpha(x,y), t_{\sigma\alpha}(x,y))$, it follows that $t_\alpha(x,y) = x$ and $t_{\sigma\alpha}(x,y) = y$ in $\mathbf{F}''$, and therefore in $\mathbf{F}$. This implies that α is the automorphism of $\mathbf{F}$ defined on generators by $(x,y) \mapsto (x,y)$, hence $\alpha = 1$ in $\mathrm{Aut}(\mathbf{F}) = G$.

We now explain why κ is surjective. Arguing as in the first paragraph, there are binary terms $\{s_\gamma \mid \gamma \in \mathrm{Aut}(\mathbf{F}'')\}$ satisfying equations like those in Definition 2.1. Since these equations hold in $\mathbf{F}''$, they hold also in $\mathbf{F}$. For each $\gamma \in \mathrm{Aut}(\mathbf{F}'')$ define an endomorphism ξ_γ of $\mathbf{F}$ on the generators by $(x,y) \mapsto (s_\gamma(x,y), s_{\sigma\gamma}(x,y))$. The equations satisfied by the s_γ's imply that $\xi_\gamma\xi_{\gamma^{-1}} = \xi_{\gamma^{-1}}\xi_\gamma = 1$ in $\mathrm{End}(\mathbf{F})$. This implies that $\xi_\gamma \in \mathrm{Aut}(\mathbf{F}) = G$ for any $\gamma \in \mathrm{Aut}(\mathbf{F}'')$. Since $\kappa(\xi_\gamma)$ is the automorphism of $\mathbf{F}''$ defined on generators by

$$(x,y) \mapsto (x\xi_\gamma, y\xi_\gamma) = (s_\gamma(x,y), s_{\sigma\gamma}(x,y)),$$

we get that $\kappa(\xi_\gamma) = \gamma$ for any $\gamma \in \mathrm{Aut}(\mathbf{F}'')$. □

One of the viewpoints that we will take in this paper is that the equations of Definition 2.1 constitute a "Maltsev condition" (cf. [3]), which we will call the $[G_\sigma]$-*Maltsev condition*. This viewpoint is not entirely adequate for our purposes, since this Maltsev condition defines the class of varieties $\mathcal{V}$ for which there is a pointed group homomorphism $G_\sigma \to \mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$, and we are interested in the more restrictive situation where there is an isomorphism $G_\sigma \cong \mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$. We will find in Theorem 2.6 that, in certain circumstances, the difference in these two situations is partially reflected by a strengthening of the $[G_\sigma]$-Maltsev condition.

Definition 2.4. Let G be a group with a designated involution σ, and let K be a subgroup of G containing σ. Then $\mathcal{V}[G_\sigma, K]$ is the subvariety of $\mathcal{V}[G_\sigma]$ defined by the equations (i)—(iii) of Definition 2.1 and

(iv) $t_\alpha(x,x) \approx x$ if $\alpha \in K$.

The equations (i)—(iv) define the $[G_\sigma, K]$-*Maltsev condition*.

Thus $\mathcal{V}[G_\sigma, K] = \mathcal{V}[G_\sigma]$ if $K = \{1, \sigma\}$, while $\mathcal{V}[G_\sigma, G]$ is the largest idempotent subvariety of $\mathcal{V}[G_\sigma]$.

Definition 2.5. Let G be a finite group with a designated involution σ, and let $C = C_G(\sigma)$ be the centralizer of σ. A subgroup $K \leq C$ is a *nucleus of* G_σ if there is a subgroup $H \leq G$ and an endomorphism ρ of C such that

(i) $H \cap H^\sigma = [H, H^\sigma] = \{1\}$, where $H^\sigma := \sigma H \sigma$,
(ii) ρ is a retraction of C onto $C \cap (HH^\sigma)$,
(iii) $\rho(\sigma) = 1$, and

(iv) $K = \ker(\rho)$.

In particular, a nucleus of G_σ is a normal subgroup of $C_G(\sigma)$ containing σ. ($C_G(\sigma)$ is itself a nucleus, arising when $H = \{1\}$ and ρ is constant.)

Theorem 2.6. *Let G_σ be a group with a designated involution σ. If $\mathbf{F}$ is the 2-generated free algebra in a subvariety $\mathcal{V} \leq \mathcal{V}[G_\sigma]$, and the canonical homomorphism $\kappa\colon G_\sigma \to Aut(\mathbf{F})_\sigma$ is an isomorphism, then $\mathbf{F}$ must lie in $\mathcal{V}[G_\sigma, K]$ for some nucleus K of G_σ.*

Proof. The assumption that κ is an isomorphism implies that every automorphism of $\mathbf{F}$ has the form $\kappa(\alpha) = \varepsilon_\alpha$ for some uniquely determined element $\alpha \in G$. (Recall that ε_α is the automorphism of $\mathbf{F}$ defined on generators by $(x, y) \mapsto (t_\alpha(x, y), t_{\sigma\alpha}(x, y))$.)

Call a unary term u *invertible* if there is another unary term v such that $u(v(x)) \approx v(u(x)) \approx x$ holds in $\mathcal{V}$.

Let $\mathbf{E} = \mathbf{F}_\mathcal{V}(x)$ be the free algebra on one generator in $\mathcal{V}$. For each $\gamma \in \mathrm{Aut}(\mathbf{E})$ let s_γ be a unary term such that $x\gamma = s_\gamma(x)$. Then $s_1(x) = x$ and

$$s_{\gamma\delta}(x) = x\gamma\delta = s_\gamma(x)\delta = s_\gamma(x\delta) = s_\gamma(s_\delta(x)), \tag{2.1}$$

so $s_1(x) \approx x$ and $s_{\gamma\delta}(x) \approx s_\gamma(s_\delta(x))$ hold in $\mathcal{V}$. In particular, $s_\gamma(s_{\gamma^{-1}}(x)) \approx s_{\gamma^{-1}}(s_\gamma(x)) \approx x$, which shows that s_γ is invertible for every $\gamma \in \mathrm{Aut}(\mathbf{E})$. Conversely, if $u(x)$ is invertible with inverse $v(x)$, then $x \mapsto u(x)$ defines an automorphism with inverse defined by $x \mapsto v(x)$. Indeed, the assignment $\gamma \mapsto s_\gamma$ is an isomorphism from $\mathrm{Aut}(\mathbf{E})$ to the group of $\mathcal{V}$-equivalence classes of invertible terms.

If $\alpha \in C_G(\sigma)$, then

$$t_{\sigma\alpha}(x, y) = t_{\alpha\sigma}(x, y) = t_\alpha(y, x), \tag{2.2}$$

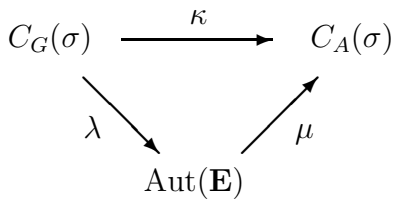
where the last equality is from the remark following Definition 2.1. Thus $x = x\alpha^{-1}\alpha = t_{\alpha^{-1}}(t_\alpha(x, y), t_\alpha(y, x))$ and $x = x\alpha\alpha^{-1} = t_\alpha(t_{\alpha^{-1}}(x, y), t_{\alpha^{-1}}(y, x))$. If we apply the endomorphism that is defined on generators by $(x, y) \mapsto (x, x)$ to these equalities we obtain that for $s_\alpha(x) := t_\alpha(x, x)$ and $s_{\alpha^{-1}}(x) := t_{\alpha^{-1}}(x, x)$ we have $s_{\alpha^{-1}}(s_\alpha(x)) = s_\alpha(s_{\alpha^{-1}}(x)) = x$. Hence $t_\alpha(x, x)$ is an invertible unary term for every $\alpha \in C_G(\sigma)$. This proves that there is a function $\lambda\colon C_G(\sigma) \to \mathrm{Aut}(\mathbf{E})$ defined by $\alpha \mapsto \gamma$ if $t_\alpha(x, x) = s_\gamma(x)$ in $\mathbf{E}$. From Definition 2.1 (iii) and line (2.2) we have

$$t_{\alpha\beta}(x, x) = t_\alpha(t_\beta(x, x), t_\beta(x, x)),$$

so λ is a group homomorphism.

Let H be the subgroup of $A := \mathrm{Aut}(\mathbf{F})$ consisting of all automorphisms defined on the generators by $(x, y) \mapsto (s_\gamma(x), y)$ for some $\gamma \in \mathrm{Aut}(\mathbf{E})$. It can be checked that H^σ consists of all automorphisms defined on the generators by $(x, y) \mapsto (x, s_\gamma(y))$ and $C_A(\sigma) \cap (HH^\sigma)$ consists of all automorphisms of $\mathbf{F}$ defined on the generators by $(x, y) \mapsto (s_\gamma(x), s_\gamma(y))$ for some $\gamma \in \mathrm{Aut}(\mathbf{E})$. It follows from line (2.1) that the function $\mu\colon \mathrm{Aut}(\mathbf{E}) \to C_A(\sigma)$ that assigns to γ the automorphism defined on

generators by $(x, y) \mapsto (s_\gamma(x), s_\gamma(y))$, is a group homomorphism. This gives us a (noncommuting) triangle of homomorphisms



where κ is an isomorphism.

Claim 2.7. *$\lambda \circ \kappa^{-1} \circ \mu$ is the identity on Aut($\mathbf{E}$).*

Proof. If $\gamma \in \mathrm{Aut}(\mathbf{E})$, then $\mu(\gamma) \in \mathrm{Aut}(\mathbf{F})$ equals $\kappa(\alpha) = \varepsilon_\alpha$ for some uniquely determined $\alpha \in G$. Since $\mu(\gamma)$ is defined on generators by $(x, y) \mapsto (s_\gamma(x), s_\gamma(y))$ and ε_α is defined on generators by $(x, y) \mapsto (t_\alpha(x, y), t_{\sigma\alpha}(x, y))$, it follows that $t_\alpha(x, y) \approx s_\gamma(x)$ holds in $\mathbf{F}$. Therefore $t_\alpha(x, x) = s_\gamma(x)$ in $\mathbf{F}$, so $\lambda(\alpha) = \gamma$. This shows that $\gamma = \lambda(\alpha) = \lambda(\kappa^{-1}(\mu(\gamma)))$ for any $\gamma \in \mathrm{Aut}(\mathbf{E})$. □

It follows from Claim 2.7 that the function $\rho := \mu \circ \lambda \circ \kappa^{-1}$ is a retraction of $C_A(\sigma)$ onto $C_A(\sigma) \cap (HH^\sigma)$. The retraction ρ is the one that takes an automorphism $\varepsilon_\alpha \in C_A(\sigma)$ that is defined on the generators by

$$(x, y) \mapsto (t_\alpha(x, y), t_{\sigma\alpha}(x, y)) = (t_\alpha(x, y), t_\alpha(y, x))$$

to the automorphism that is defined on the generators by $(x, y) \mapsto (t_\alpha(x, x), t_\alpha(y, y))$. In particular, since σ is defined on the generators by $(x, y) \mapsto (y, x)$, $\rho(\varepsilon_\sigma)$ is defined on the generators by $(x, y) \mapsto (x, y)$; i.e., $\rho(\sigma) = 1$. According to Definition 2.5, this forces $K_A := \ker(\rho)$ to be a nucleus of $\mathrm{Aut}(\mathbf{F})_\sigma$.

Choose any $\alpha \in G_\sigma$ such that $\kappa(\alpha) = \varepsilon_\alpha \in K_A$ ($\subseteq C_A(\sigma)$). Then $\alpha \in C_G(\sigma)$, so ε_α is defined on the generators by $(x, y) \mapsto (t_\alpha(x, y), t_\alpha(y, x))$. From the previous paragraph, we get that $\rho(\varepsilon_\alpha)$ is defined on the generators by $(x, y) \mapsto (t_\alpha(x, x), t_\alpha(y, y))$. But $\rho(\varepsilon_\alpha) = 1$ if $\varepsilon_\alpha \in K_A$, so $(t_\alpha(x, x), t_\alpha(y, y)) = (x, y)$ in $\mathbf{F}$. This proves that $t_\alpha(x, x) \approx x$ is an equation that holds in $\mathbf{F}$ for every $\alpha \in G_\sigma$ for which $\kappa(\alpha) \in K_A$. Since κ is an isomorphism and nuclei are intrinsically-defined subgroups, $K := \kappa^{-1}(K_A)$ is a nucleus of G_σ. Now $\alpha \in K$ if and only if $\kappa(\alpha) \in K_A$, and these conditions imply that $t_\alpha(x, x) \approx x$ holds in $\mathbf{F}$. This shows that $\mathbf{F} \in \mathcal{V}[G_\sigma, K]$ for some nucleus K. □

It can be shown that if G_σ has a retraction onto $\langle\sigma\rangle$, then the $[G_\sigma]$-Maltsev condition is trivial, i.e., it is satisfied by every variety. (If $\rho\colon G_\sigma \to \langle\sigma\rangle$ is a retraction, then the assignment $t_\alpha(x, y) \mapsto x$ if $\rho(\alpha) = 1$ and $t_\alpha(x, y) \mapsto y$ if $\rho(\alpha) = \sigma$ defines an interpretation of $\mathcal{V}[G_\sigma]$ into any variety. Indeed, this even shows that the stronger $[G_\sigma, G]$-Maltsev condition is trivial.) Even if G_σ has no retraction onto $\langle\sigma\rangle$, then the $[G_\sigma]$-Maltsev condition is nearly trivial: if $|G| = 2k$, then it can be shown that

the $[G_\sigma]$-Maltsev condition is satisfied by the k-th matrix power of any variety. It is therefore surprising to find that if a subgroup K of G_σ containing σ has no retraction onto $\langle\sigma\rangle$, then the $[G_\sigma, K]$-Maltsev condition implies congruence permutability, which is equivalent to a very strong Maltsev condition. We begin this proof now.

Lemma 2.8. *Let $\mathcal{V}$ be an idempotent variety that is not congruence permutable. If $\mathbf{F} = \mathbf{F}_\mathcal{V}(x, y)$ is the 2-generated free algebra in $\mathcal{V}$, then $\mathbf{F}$ has subuniverses U and V such that*

(1) *$x \in U$, $x \in V$,*
(2) *$y \notin U$, $y \notin V$, and*
(3) *$(U \times F) \cup (F \times V)$ is a subuniverse of $\mathbf{F} \times \mathbf{F}$.*

Proof. Let S_0 be the subuniverse of $\mathbf{F}^2$ that is generated by $\{(x, y), (x, x), (y, x)\}$. The pair (y, y) is in this subuniverse if and only if $\mathcal{V}$ has a ternary term p such that $p((x, y), (x, x), (y, x)) = (y, y)$ in $\mathbf{F}^2$, or equivalently $p(x, x, y) = y$ and $p(y, x, x) = y$ in $\mathbf{F}$. Since $\mathbf{F} = \mathbf{F}_\mathcal{V}(x, y)$ is free, this happens if and only if p is a Maltsev term for $\mathcal{V}$. Since we have assumed that $\mathcal{V}$ is not congruence permutable, $(y, y) \notin S_0$.

Since $\mathcal{V}$ is idempotent, $\mathrm{Sg}^{\mathbf{F^2}}(\{(x, x), (x, y)\}) = \{x\} \times \mathrm{Sg}^{\mathbf{F}}(\{x, y\}) = \{x\} \times F$. Thus, if we take $U_0 = \{x\}$, then $\mathrm{Sg}^{\mathbf{F^2}}((U_0 \times F) \cup \{(y, x)\}) = S_0$, which implies that $(y, y) \notin \mathrm{Sg}^{\mathbf{F^2}}((U_0 \times F) \cup \{(y, x)\})$. Using Zorn's lemma, extend U_0 to a subuniverse U of $\mathbf{F}$ that is maximal for the condition that $(y, y) \notin S := \mathrm{Sg}^{\mathbf{F^2}}((U \times F) \cup \{(y, x)\})$. Let V be the subset of F defined by the equation $\{y\} \times V = S \cap (\{y\} \times F)$. That V is a subuniverse of $\mathbf{F}$ follows from the facts that S is a subuniverse of $\mathbf{F}^2$ and $\mathcal{V}$ is idempotent. We now argue that U and V have the required properties.

The fact that $x \in U$ follows from the fact that U extends the subuniverse $U_0 = \{x\}$. That $x \in V$ follows from the fact that $(y, x) \in S \cap (\{y\} \times F) = \{y\} \times V$. If $y \in U$, then U would contain the generators of $\mathbf{F}$, which would lead to the contradiction that $F \times F = U \times F \subseteq S \subsetneq F \times F$. If $y \in V$, then $(y, y) \in (\{y\} \times V) \subseteq S$, which is false. These arguments show that (1) and (2) hold.

To prove that (3) holds, it will suffice to show that $(U \times F) \cup (F \times V) = S$, since S is a subuniverse of $\mathbf{F}^2$. It follows from the definition of S that $U \times F \subseteq S$. If $v \in V$, then the definition of V implies that $(y, v) \in S$. Moreover $(x, v) \in S$ since $U \times F \subseteq S$ and $x \in U$. Thus, if $v \in V$, then S contains $\mathrm{Sg}^{\mathbf{F^2}}(\{(x, v), (y, v)\}) = F \times \{v\}$. This is enough to show that $F \times V \subseteq S$, hence $(U \times F) \cup (F \times V) \subseteq S$.

To establish the reverse inclusion, assume that $(p, q) \in S - (U \times F)$. Then $p \notin U$, so if $U' := \mathrm{Sg}^{\mathbf{F}}(U \cup \{p\})$ then the maximality of U implies that

$$(y, y) \in \mathrm{Sg}^{\mathbf{F^2}}((U' \times F) \cup \{(y, x)\}). \tag{2.3}$$

Claim 2.9. *$Sg^{\mathbf{F^2}}((U' \times F) \cup \{(y, x)\})$ is generated by*

$$X := (U \times \{y\}) \cup \{(p, y), (x, x), (y, x)\}.$$

Proof. Since $(U \times \{y\}) \cup \{(p,y),(x,x)\} \subseteq U' \times F$, the subalgebra generated by X is contained in $\mathrm{Sg}^{\mathbf{F^2}}((U' \times F) \cup \{(y,x)\})$. To establish equality it will suffice to show that $U' \times F \subseteq \mathrm{Sg}^{\mathbf{F^2}}(X)$. We have $U' \times \{y\} = \mathrm{Sg}^{\mathbf{F^2}}((U \times \{y\}) \cup \{(p,y)\}) \subseteq \mathrm{Sg}^{\mathbf{F^2}}(X)$, and $U' \times \{x\} \subseteq F \times \{x\} = \mathrm{Sg}^{\mathbf{F^2}}(\{(x,x),(y,x)\}) \subseteq \mathrm{Sg}^{\mathbf{F^2}}(X)$. Thus, for every $u \in U'$ we have $(u,x),(u,y) \in \mathrm{Sg}^{\mathbf{F^2}}(X)$, so $\{u\} \times F = \mathrm{Sg}^{\mathbf{F^2}}(\{(u,x),(u,y)\} \subseteq \mathrm{Sg}^{\mathbf{F^2}}(X)$. This shows that $U' \times F \subseteq \mathrm{Sg}^{\mathbf{F^2}}(X)$, ands completes the proof of the claim. □

Applying Claim 2.9 to line (2.3) leads to the conclusion that there is a term r such that $(y,y) = r((p,y),(x,x),(y,x),(u_1,y),\ldots,(u_k,y))$ where all u_i are from U. Writing the coordinate equations separately yields that $y = r(p,x,y,u_1,\ldots,u_k)$ and $y = r(y,x,x,y,\ldots,y)$ in $\mathbf{F}$. Applying the endomorphism of $\mathbf{F}$ determined on the generators by $x \mapsto x$ and $y \mapsto q$ to the second of these equations, we derive that $q = r(q,x,x,q,\ldots,q)$. Recombining this with the first of the equations yields that

$$(y,q) = r((p,q),(x,x),(y,x),(u_1,q),\ldots,(u_k,q))$$

in $\mathbf{F}^2$. Since all of the pairs $(p,q),(x,x),(y,x),(u_1,q),\ldots,(u_k,q)$ belong to S, it follows that the pair (y,q) is in S. This shows that $(y,q) \in S \cap (\{y\} \times F) = \{y\} \times V$, and therefore that $q \in V$. Altogether this shows that if $(p,q) \in S - (U \times F)$, then $(p,q) \in F \times V$. Hence $S \subseteq (U \times F) \cup (F \times V)$. □

Theorem 2.10. *Let K be a finite group with a designated involution σ. If K does not have a retraction onto $\langle\sigma\rangle$, then $\mathcal{V}[K_\sigma, K]$ is congruence permutable.*

Proof. We prove the contrapositive. Assume that $\mathcal{V}[K_\sigma, K]$ is not congruence permutable. Since $\mathcal{V}[K_\sigma, K]$ is idempotent, Lemma 2.8 guarantees that the algebra $\mathbf{F} \in \mathcal{V}[K_\sigma, K]$ that is freely generated by $\{x,y\}$ has subuniverses U and V, each containing x but not y, such that $(U \times F) \cup (F \times V)$ is a subuniverse of $\mathbf{F}^2$.

Claim 2.11. *Let $\mathcal{L} := \{\alpha \in K \mid t_\alpha(U,F) \subseteq U\}$ and $\mathcal{R} := \{\alpha \in K \mid t_\alpha(F,V) \subseteq V\}$.*

(i) *$1 \in \mathcal{L} - \mathcal{R}$ and $\sigma \in \mathcal{R} - \mathcal{L}$.*
(ii) *$\mathcal{L}$ is a subgroup of K.*
(iii) *$\mathcal{R}\sigma$ is a subgroup of K.*

Proof. For (i), $t_1(x,y) = x$ and $t_\sigma(x,y) = y$, so $t_1(U,F) \subseteq U$ and $t_\sigma(F,V) \subseteq V$, and therefore $1 \in \mathcal{L}$ and $\sigma \in \mathcal{R}$. Since $x \in V$ and $y = t_1(y,x) \notin V$ we get that $1 \notin \mathcal{R}$. Since $x \in U$ and $y = t_\sigma(x,y) \notin U$ we get $\sigma \notin \mathcal{L}$.

For (ii), assume that $\alpha, \beta \in \mathcal{L}$. Then by Definition 2.1 (iii)

$$t_{\alpha\beta}(U,F) \subseteq t_\alpha(t_\beta(U,F), t_{\sigma\beta}(U,F)) \subseteq t_\alpha(U,F) \subseteq U\,,$$

so $\alpha\beta \in \mathcal{L}$.

For (iii), assume that $\alpha, \beta \in \mathcal{R}\sigma$. Then $\alpha\sigma, \beta\sigma \in \mathcal{R}$, so

$$t_{\alpha\beta\sigma}(F,V) \subseteq t_\alpha(t_{\beta\sigma}(F,V), t_{\sigma\beta\sigma}(F,V)) \subseteq t_\alpha(V,F) = t_{\alpha\sigma}(F,V) \subseteq V\,.$$

This shows that $\alpha\beta\sigma \in \mathcal{R}$, so $\alpha\beta \in \mathcal{R}\sigma$. This proves the claim. □

Claim 2.12. $K = \mathcal{L} \cup \mathcal{R}$.

Proof. Assume instead that K has an element $\alpha \notin \mathcal{L} \cup \mathcal{R}$. Then there exist elements $p \in U, q, r \in F$ and $s \in V$ such that $t_\alpha(p, q) \notin U$ and $t_\alpha(r, s) \notin V$. Then $(p, r) \in U \times F$ and $(q, s) \in F \times V$, but

$$t_\alpha((p, r), (q, s)) \notin (U \times F) \cup (F \times V)\,.$$

This contradicts the fact that $(U \times F) \cup (F \times V)$ is a subuniverse of $\mathbf{F}^2$, so the claim is proved. □

Now we complete the proof of the theorem. We have shown in Claims 2.11 and 2.12 that K is the union of a proper subgroup $\mathcal{L}$ and a coset $\mathcal{R}$ of a proper subgroup $\mathcal{R}\sigma$. This implies that $\mathcal{L} = \mathcal{R}\sigma$ is a subgroup of index 2. Since $\sigma \notin \mathcal{L}$, this subgroup is the kernel of a retraction onto $\langle\sigma\rangle$. The theorem is proved. □

Corollary 2.13. *Let G be a finite group with a designated involution σ, and let K be a subgroup of G containing σ. If K does not have a retraction onto $\langle\sigma\rangle$, then $\mathcal{V}[G_\sigma, K]$ is congruence permutable.*

Proof. $\mathcal{V}[G_\sigma, K]$ satisfies the $[K_\sigma, K]$-Maltsev condition. Since K has no retraction onto $\langle\sigma\rangle$, it follows from Theorem 2.10 that this Maltsev condition implies congruence permutability. □

Example 2.14. Let G be a finite group with a designated involution σ. Suppose that there is a $\gamma \in G$ such that $\gamma^2 = \sigma$. Let $K = \langle\gamma\rangle = \{1, \gamma, \gamma^2, \gamma^3\}$. Since K does not have a retraction onto $\langle\sigma\rangle = \{1, \gamma^2\}$, Corollary 2.13 implies that $\mathcal{V}[G_\sigma, K]$ is congruence permutable. In this special case, a computer search has located the shortest Maltsev term for $\mathcal{V}[G_\sigma, K]$, it is

$$p(x, y, z) = [((y(yx))((yx)z))((yz)x)][(x(zy))((z(xy))((xy)y))]\,,$$

where the product xy is shorthand for $t_\gamma(x, y)$.

Corollary 2.15. *Let G be a finite group with a designated involution σ. Assume that no nucleus of G_σ has a retraction onto $\langle\sigma\rangle$. If $G_\sigma \cong Aut(\mathbf{F}_\mathcal{W}(x, y))_\sigma$, then $\mathcal{W}$ is a congruence permutable variety.*

Proof. By Theorem 2.3, some reduct $\mathcal{V}$ of $\mathcal{W}$ is a subvariety of $\mathcal{V}[G_\sigma]$ and has the property that the canonical homomorphism $\kappa\colon G_\sigma \to \mathrm{Aut}(\mathbf{F}_\mathcal{V}(x, y))_\sigma$ is an isomorphism. If $\mathcal{V}$ is congruence permutable, then $\mathcal{W}$ will be, so it suffices to consider only the case where $\mathcal{W} = \mathcal{V}$.

By Theorem 2.6, $\mathbf{F}_\mathcal{V}(x, y) \in \mathcal{V}[G_\sigma, K]$ where K is some nucleus of G_σ. Since no nucleus has a retraction onto $\langle\sigma\rangle$, it follows from Corollary 2.13 that $\mathcal{V}[G_\sigma, K]$, and hence $\mathcal{V}$, is congruence permutable. □

If G is a group with a designated involution σ, then we will say that G_σ *forces congruence permutability* if a variety $\mathcal{V}$ is congruence permutable whenever $G_\sigma \cong \mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$. The subscript σ will be dropped if G_σ forces congruence permutability for every involution $\sigma \in G$.

Corollary 2.15 shows that a finite pointed group G_σ forces congruence permutability if no nucleus K of G_σ has a retraction onto $\langle\sigma\rangle$. If $\rho\colon K \to \langle\sigma\rangle$ were a retraction, then $N := \ker(\rho)$ would be a normal complement of $\langle\sigma\rangle$ in K. Since $K \leq C_G(\sigma)$ we have $\sigma \in Z(K)$, so $K \cong \langle\sigma\rangle \times N$. Thus, Corollary 2.15 may be reworded to state that if $\langle\sigma\rangle$ is not a direct factor of any nucleus, then G_σ forces congruence permutability.

For later reference we identify some groups that do not force congruence permutability.

Theorem 2.16. *Let G be a finite group with designated involution σ. If either*

1. *there is a retraction $\rho\colon G \twoheadrightarrow \langle\sigma\rangle$, or*
2. *G is the symmetric group S_{2k} and σ is a fixed point free involution,*

then G_σ does not force congruence permutability. In both cases, there is a finitely generated variety $\mathcal{V}$ satisfying no nontrivial idempotent Maltsev condition such that such that $G_\sigma \cong Aut(\mathbf{F}_\mathcal{V}(x,y))_\sigma$. If (1) holds, $\mathcal{V}$ can be taken to be idempotent.

Proof. The argument for case (1) is based on Gould's second construction in [4] and his results in [5]. His work shows that if $\mathbf{F}$ is the algebra with universe G and basic operations t_α, $\alpha \in G$, defined by

$$t_\alpha(\beta,\gamma) := \begin{cases} \alpha\beta & \text{if } \gamma = \sigma\beta, \\ \beta & \text{if } \gamma \neq \sigma\beta \text{ and } \rho(\alpha) = 1, \\ \gamma & \text{if } \gamma \neq \sigma\beta \text{ and } \rho(\alpha) = \sigma, \end{cases}$$

then $\mathbf{F}$ is free over $\{1,\sigma\}$, and the automorphisms of $\mathbf{F}$ are precisely the the right multiplications $R_\alpha\colon x \mapsto x\alpha$. That is, the right regular representation of G is an isomorphism of the pointed group G_σ onto $\mathrm{Aut}(\mathbf{F})_\sigma$.

It is easy to check that each t_α is an idempotent operation. Moreover, $\theta := \ker(\rho)$ is compatible with each t_α, hence is a congruence of $\mathbf{F}$. The quotient $\mathbf{F}/\theta$ is a 2-element algebra satisfying $t_\alpha(x,y) \approx x$ when $\rho(\alpha) = 1$ and $t_\alpha(x,y) \approx y$ when $\rho(\alpha) = \sigma$. Thus, $\mathbf{F}/\theta$ is term equivalent to the 2-element set, implying that $\mathbf{F}$ belongs to no variety satisfying a nontrivial idempotent Maltsev condition. This establishes all claims concerning case (1).

Concerning case (2), the k-th matrix power of the variety of sets is a finitely generated variety that satisfies no nontrivial idempotent Maltsev condition, and for this variety $\mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma \cong (S_{2k})_\sigma$ where σ is a fixed point free involution. □

3. Examples

The following examples illustrate that the hypothesis of Corollary 2.15 can be satisfied. The results of these examples will be used in Section 6.

Notation that will be fixed throughout these examples include that G is a finite group with designated involution σ, $C := C_G(\sigma)$ is the centralizer of σ, $H \leq G$ is a subgroup of G satisfying $H \cap H^\sigma = [H, H^\sigma] = 1$, ρ is a retraction of C onto $C \cap (HH^\sigma)$ satisfying $\rho(\sigma) = 1$, and K is the kernel of ρ. Our task will be to show that for H, ρ and K related in this way, there is no retraction of K onto $\langle\sigma\rangle$. Observe that the conditions on H imply that the function $(h, k) \mapsto hk^\sigma$ is an isomorphism from $H \times H$ to HH^σ, and under this isomorphism the diagonal of $H \times H$ corresponds to $C \cap (HH^\sigma)$ (so $C \cap (HH^\sigma) \cong H$).

In these examples we will usually choose σ to be a 2-*central involution* of G, by which we mean an involution that belongs to the center of some 2-Sylow subgroup of G (equivalently C contains a 2-Sylow subgroup of G, equivalently C has odd index, equivalently σ has an odd number of conjugates). There is no theoretical reason to prefer 2-central involutions — there exist varieties where the automorphism that interchanges the free generators of $\mathbf{F}_\mathcal{V}(x, y)$ is 2-central in $\mathrm{Aut}(\mathbf{F}_\mathcal{V}(x, y))$, and there exist varieties where it is not — but some of the arguments below are harder or false when σ is not 2-central.

Example 3.1. Assume that C is a normal Hall subgroup of G. We will argue that G_σ has exactly one nucleus, namely $K = C$.

The fact that $H \cap H^\sigma = \{1\}$ forces $C \cap H = \{1\}$. Since C is normal, $|H|$ divides $[G : C]$. Since ρ maps C onto $C \cap (HH^\sigma) \cong H$, $|H|$ also divides $|C|$. From $\gcd([G : C], |C|) = 1$ it follows that $|H| = 1$, so ρ is constant. Hence $K = C$ is the only nucleus. From Corollary 2.15, if C is a normal Hall subgroup and C has no retraction onto $\langle\sigma\rangle$, then G_σ forces congruence permutability.

Example 3.2. As a special case of Example 3.1, let σ belong to the center of G. Then $K = C = G$ is the only nucleus of G_σ. If G has no retraction onto $\langle\sigma\rangle$, then G_σ forces congruence permutability.

Example 3.3. The 2-groups that contain a unique involution are the cyclic 2-groups or the generalized quaternion groups. If the 2-Sylow subgroups of G are of this type, then G contains exactly one conjugacy class of involutions, and all involutions are 2-central. We will argue that every nucleus of G_σ contains every 2-Sylow subgroup containing σ.

If σ and τ are commuting involutions of G, then the subgroup they generate can be extended to a 2-Sylow subgroup of G. But each 2-Sylow subgroup of G contains a unique involution, so $\sigma = \tau$. This proves that σ is the unique involution of C. Therefore $\sigma \in P_2 \iff P_2 \subseteq C$ when P_2 is a 2-Sylow subgroup of G.

Since ρ is a retraction and $\rho(\sigma) = 1$, it must be that $\sigma \notin \mathrm{im}(\rho)$. Since $\mathrm{im}(\rho) \subseteq C$, the conclusion of the previous paragraph implies that $\mathrm{im}(\rho)$ contains no involution, hence $\mathrm{im}(\rho)$ has odd order. Thus every 2-Sylow subgroup of C is contained in $\ker(\rho) = K$. Since a 2-Sylow subgroup of C is just a 2-Sylow subgroup of G containing σ, every nucleus K of G_σ contains every 2-Sylow subgroup that contains σ.

In particular, if some nucleus had a retraction onto $\langle\sigma\rangle$, then the retraction could be restricted to one of a 2-Sylow subgroup onto $\langle\sigma\rangle$. But if the 2-Sylow subgroups of G are generalized quaternion or cyclic of order at least 4, then there can be no such retraction. In these cases, G forces congruence permutability.

Example 3.4. Let $G = \mathrm{SL}(n, q)$ where $n > 1$ and q is an odd prime power greater than 3. We will show that for any involution $\sigma \in G$ and any nucleus K of G_σ, the subgroup K has no retraction onto $\langle\sigma\rangle$. This will prove that $\mathrm{SL}(n, q)$ forces congruence permutability.

To show that G_σ forces congruence permutability for any σ, it suffices to show it for one from each conjugacy class. Therefore, we may assume that σ is equal to a block diagonal matrix of the form

$$\begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}.$$

Here the upper left block is the negative of the $m \times m$ identity matrix where m is a positive even integer, the bottom right block is the $\ell \times \ell$ identity matrix where $\ell \geq 0$, and $m + \ell = n$. Every involution in $\mathrm{SL}(n, q)$ is conjugate to such a matrix.

A computation shows that C is the group of all invertible block diagonal matrices of the form

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where $\det(A) = \det(B)^{-1}$. This representation makes it clear that C is isomorphic to the subdirect product of $\mathrm{GL}(m, q)$ and $\mathrm{GL}(\ell, q)$ consisting of all pairs (A, B) with $\det(A) = \det(B)^{-1}$. We will use this more compact representation for C (so, for example, $\sigma = (-I, I)$).

We assert that every nucleus of G_σ contains every matrix with the representation (A, I), where $A \in \mathrm{SL}(m, q)$. According to Definition 2.5, to prove this it will suffice to show that if ρ is a retraction of C satisfying $\rho(\sigma) = r((-I, I)) = (I, I) = 1$, then $\rho(\mathrm{SL}(m, q) \times \{I\}) = \{(I, I)\}$.

Fix any retraction ρ of C satisfying $\rho((-I, I)) = (I, I)$. This retraction must map the fully invariant subgroup $[C, C] = \mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q)$ into itself. We use the same symbol ρ to denote the restriction of ρ to $[C, C]$ or any other subgroup of G.

Claim 3.5. *A nonconstant endomorphism of a special linear group* $\mathrm{SL}(k, q)$*, with* $k > 1$*,* q *odd, and* $(k, q) \neq (2, 3)$ *is an automorphism.*

Proof. When $k > 1$, q is odd, and $(k, q) \neq (2, 3)$, the group $\mathrm{SL}(k, q)$ is a *quasisimple* group (a perfect group G such that $G/Z(G)$ is simple). Suppose that $\varepsilon\colon G \to G$ is a nonconstant endomorphism. The kernel of ε is a proper normal subgroup of G, hence lies in the center of G. Therefore $\mathrm{im}(\varepsilon) = H$ is also quasisimple and has the same simple factor as G. This means that $HZ(G)$ has the same composition factors as G, so $HZ(G) = G$. This forces H to be normal in G, and not in the center, so $G = H = \mathrm{im}(\varepsilon)$. Since G is finite, ε is an automorphism. □

Claim 3.6. *The composite homomorphism*

$$\rho_1\colon \mathrm{SL}(m, q) \times \{I\} \xrightarrow{\rho} \mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q) \xrightarrow{\pi_1} \mathrm{SL}(m, q)$$

is constant.

Proof. Let τ be the isomorphism $\mathrm{SL}(m, q) \to \mathrm{SL}(m, q) \times \{I\}\colon A \mapsto (A, I)$. Then $\rho_1 \circ \tau$ is an endomorphism of $\mathrm{SL}(m, q)$ that is not an automorphism (since $\rho_1 \circ \tau(-I) = \rho_1(\sigma) = I$). By Claim 3.5, $\rho_1 \circ \tau$ is constant. Since τ is an isomorphism, ρ_1 is constant. □

Claim 3.7. *If* $\mathrm{SL}(m, q) \times \{I\} \not\subseteq \ker(\rho)$*, then*

$$\rho(\mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q)) = \{I\} \times \mathrm{SL}(\ell, q).$$

Proof. Let ρ_2 be the composite $\mathrm{SL}(m, q) \times \{I\} \xrightarrow{\rho} \mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q) \xrightarrow{\pi_2} \mathrm{SL}(\ell, q)$, so $\rho(x) = (\rho_1(x), \rho_2(x))$. If $\mathrm{SL}(m, q) \times \{I\} \not\subseteq \ker(\rho)$, then since ρ_1 is constant we have

$$\rho(\mathrm{SL}(m, q) \times \{I\}) = \{I\} \times \rho_2(\mathrm{SL}(m, q) \times \{I\})$$

where $M := \rho_2(\mathrm{SL}(m, q) \times \{I\})$ is a nontrivial subgroup of $\mathrm{SL}(\ell, q)$. A nontrivial homomorphic image of a perfect group is perfect, hence nonsolvable, so M is a nonsolvable subgroup of $\mathrm{SL}(\ell, q)$. This proves that $\ell > 1$. Since $\{I\} \times M$ is in the image of the retraction ρ, it follows that ρ maps the normal subgroup generated by $\{1\} \times M$ into itself. That subgroup can only be $\{1\} \times \mathrm{SL}(\ell, q)$, so ρ restricts to a retraction of $\{1\} \times \mathrm{SL}(\ell, q)$ onto a subgroup containing $\{1\} \times M$. But since $\ell > 1$, it follows from Claim 3.5 that any retraction of $\{1\} \times \mathrm{SL}(\ell, q)$ ($\cong \mathrm{SL}(\ell, k)$) is constant or the identity. The former possibility is contradicted by the nontriviality of M, forcing $\rho(x) = x$ on $\{I\} \times \mathrm{SL}(\ell, q)$.

We have shown that ρ maps $\mathrm{SL}(m, q) \times \{I\}$ to $\{1\} \times M \subseteq \{1\} \times \mathrm{SL}(\ell, q)$, and that it maps $\{I\} \times \mathrm{SL}(\ell, q)$ onto itself. Thus ρ maps a generating set for $\mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q)$ into $\{I\} \times \mathrm{SL}(\ell, q)$ with a subset of generators mapping onto $\{I\} \times \mathrm{SL}(\ell, q)$. It follows that $\rho(\mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q)) = \{I\} \times \mathrm{SL}(\ell, q)$ when $\mathrm{SL}(m, q) \times \{I\} \not\subseteq \ker(\rho)$. □

Claim 3.8. $\mathrm{SL}(m, q) \times \{I\} \subseteq \ker(\rho)$.

Proof. If this is not so, then Claim 3.7 yields

$$\rho(\mathrm{SL}(m, q) \times \mathrm{SL}(\ell, q)) = \{I\} \times \mathrm{SL}(\ell, q).$$

Then $\ker(\rho)$, restricted to the product $\mathrm{SL}(m,q)\times\mathrm{SL}(\ell,q)$, is a normal complement of $\rho(\mathrm{SL}(m,q)\times\mathrm{SL}(\ell,q)) = \{I\}\times\mathrm{SL}(\ell,q)$. But $\mathrm{SL}(m,q)\times\mathrm{SL}(\ell,q)$ is a perfect group, so any normal subgroup (in this case $\{I\}\times\mathrm{SL}(\ell,q)$) has at most one normal complement (which in this case can only be $\mathrm{SL}(m,q)\times\{1\}$). Thus $\mathrm{SL}(m,q)\times\{I\}\subseteq\ker(\rho)$, indeed. □

Claim 3.8 is exactly what we had to establish to show that any nucleus of G_σ contains $\mathrm{SL}(m,q)\times\{I\}$.

Finally, it is easy to see why no nucleus K of G_σ has a retraction onto $\langle\sigma\rangle$. Since

$$\{1,\sigma\} = \{(I,I),(-I,I)\}\subseteq\mathrm{SL}(m,q)\times\{I\}\subseteq K,$$

any retraction of K onto $\langle\sigma\rangle$ would restrict to a retraction of $\mathrm{SL}(m,q)\times\{I\}$ onto its subgroup $\{\pm I\}\times\{I\}$. This would imply the existence of a retraction of $\mathrm{SL}(m,q)$ onto $\{\pm I\}$, contradicting Claim 3.5.

Remark 3.9. The argument of the previous example applies without change to the case $q=3$ if σ has more than 2 negative eigenvalues. When $q=3$ and σ has exactly 2 negative eigenvalues, then each of the Claims is false. Nevertheless, it can be shown that the final result is true: for any involution $\sigma\in\mathrm{SL}(n,3)$, no nucleus of $\mathrm{SL}(n,3)_\sigma$ has a retraction onto $\langle\sigma\rangle$. Therefore $\mathrm{SL}(n,q)$ forces congruence permutability whenever $n>1$ and q is an odd prime power. On the other hand, we do not know if $\mathrm{SL}(n,q)$ forces congruence permutability when q is even.

By Theorem 2.16, $(S_n)_\sigma$ does not force congruence permutability when σ is a fixed point free involution. Somewhat surprisingly, the alternating group A_n, $n>4$, does force congruence permutability when σ is a fixed point free involution.

Example 3.10. If A_n contains a fixed point free involution σ, then necessarily $n=4k$ for some k and σ is a product of $2k$ disjoint transpositions. The purpose of this example is to show that $(A_n)_\sigma$ forces congruence permutability if $n=4k>4$.

Claim 3.11. *If $m>2$, then $S_m\times S_m$ is not embeddable in A_{2m}.*

Proof. If the claim is not true, then there exists a faithful action of the group $S_m\times S_m$ on a set X of size $2m$ where every group element acts as an even permutation. This may be interpreted as a faithful action of S_m on X where every element acts evenly, and where the automorphism group of the S_m-set X contains a subgroup isomorphic to S_m acting evenly. We make use of the last interpretation.

Since S_m is subdirectly irreducible and acts faithfully on X, it follows that S_m acts faithfully on a single orbit $Y\subseteq X$. It must be that Y has size at least $m=|X|/2$, since S_m has no smaller faithful action. If there is a second orbit Y' on which S_m acts faithfully, then necessarily $|Y|=|Y'|=|X|/2$. In this case, $|\operatorname{Aut}(\langle X;S_m\rangle)|\leq 2$, since an S_m-set of size m is rigid when $m>2$. Hence S_m does not embed into the automorphism group of such an S_m-set, forcing Y to be the only orbit on which S_m acts faithfully.

Let $Z = X - Y$. By the previous paragraph, $|Z| \leq |X|/2 = m$ and S_m does not act faithfully on Z. Since Y is a single orbit, and it is not isomorphic to any other orbit,

$$\mathrm{Aut}(\langle X; S_m\rangle) \cong \mathrm{Aut}(\langle Y; S_m\rangle) \times \mathrm{Aut}(\langle Z; S_m\rangle).$$

We have assumed that S_m is embeddable in $\mathrm{Aut}(\langle X; S_m\rangle)$, so since it is a subdirectly irreducible group S_m is embeddable in either $\mathrm{Aut}(\langle Y; S_m\rangle)$ or $\mathrm{Aut}(\langle Z; S_m\rangle)$. If S_m is embeddable in $\mathrm{Aut}(\langle Z; S_m\rangle)$, then the elements of Z can be permuted in at least $m!$ ways. Since $|Z| \leq m$, we conclude that $|Z| = m$ and every permutation of Z is an automorphism of the S_m-set $\langle Z; S_m\rangle$. This can only happen if Z consists of m one-element S_m-orbits. In this case, $X = Y \cup Z$ has one orbit of size m and m orbits of size one. But this is not an action of S_m on X where every element acts evenly. It must therefore be that S_m embeds in $\mathrm{Aut}(\langle Y; S_m\rangle)$. If the stabilizer of a point $y \in Y$ is H, then the fact that Y is a single orbit implies that $\mathrm{Aut}(\langle Y; S_m\rangle) \cong N_{S_m}(H)/H$. For S_m to embed in this group we must have $H = \{1\}$. But then S_m acts regularly on Y, so $|Y| = m!$. We have assumed that $m > 2$, so from $m! = |Y| \leq 2m$ we derive that $m = 3$ and $Y = X$. But this cannot be so, since when S_3 acts regularly on X the transpositions of S_3 do not act evenly on X. This contradiction completes the proof. □

To fix notation for the rest of this example, let $n = 4k$ and assume that the set of $4k$ letters on which A_{4k} acts is $X = \{1, 2, \ldots, 2k\} \cup \{\overline{1}, \ldots, \overline{2k}\}$ and that $\sigma = (1\ \overline{1})(2\ \overline{2}) \cdots (2k\ \overline{2k})$. We define $\overline{\overline{i}} = i$, so that in fact $\sigma(x) = \overline{x}$ for any $x \in X$.

For each $1 \leq i \leq 2k$, choose an element $i^* \in \{i, \overline{i}\}$. For each $\alpha \in S_{2k}$, define a permutation α^* on X by $\alpha^*(i^*) = (\alpha(i))^*$ and $\alpha^*(\overline{i}^*) = \overline{(\alpha(i))^*}$, i.e., α^* permutes each of the sets $\{1^*, \ldots, (2k)^*\}$ and $\{\overline{1^*}, \ldots, \overline{(2k)^*}\}$ the same way that α permutes $\{1, \ldots, 2k\}$. Then $\alpha^* \in A_{4k}$, α^* commutes with σ, and $S_{2k}^* := \{\alpha^* \mid \alpha \in S_{2k}\}$ is a subgroup of $C_{A_{4k}}(\sigma)$ that is isomorphic to S_{2k}.

Let C, H, ρ and K be as in the definition of a nucleus. H cannot contain a subgroup isomorphic to S_{2k}, since if it did $S_{2k} \times S_{2k} \leq H \times H \cong HH^\sigma$ would be embeddable in A_{4k}, contrary to Claim 3.11. Therefore, the diagonal $C \cap (HH^\sigma) \cong H$ of the square subgroup HH^σ does not have a subgroup isomorphic to S_{2k}. Since $C \cap (HH^\sigma) = \rho(C)$, ρ cannot restrict to a 1-1 function on S_{2k}^*, hence $K = \ker(\rho)$ contains the smallest nontrivial normal subgroup of S_{2k}^* for any choice function $*$. In particular, K contains all elements of the form $\beta := (i\ j)(k\ \ell)(\overline{i}\ \overline{j})(\overline{k}\ \overline{\ell})$ and $\gamma := (\overline{i}\ j)(k\ \ell)(i\ \overline{j})(\overline{k}\ \overline{\ell})$, and therefore all elements of the form $\beta\gamma = (i\ \overline{i})(j\ \overline{j})$. The kernel of any retraction $\rho' \colon K \to \langle\sigma\rangle$ contains $[K, K]$, so for $\delta := (i\ \overline{i})(k\ \overline{k}) \in K$ we have

$$(i\ \overline{i})(j\ \overline{j}) = [(i\ j)(k\ \ell)(\overline{i}\ \overline{j})(\overline{k}\ \overline{\ell}), (i\ \overline{i})(k\ \overline{k})] = [\beta, \delta] \in [K, K] \leq \ker(\rho').$$

But $\sigma = (1\ \overline{1})(2\ \overline{2}) \cdots (2k\ \overline{2k})$ is a product of k elements of this type, so $\sigma \in \ker(\rho')$. This contradicts the fact that ρ' is a retraction with σ in its image.

Remark 3.12. Using virtually identical arguments, it can be shown that if $n = 4k+1 \geq 9$ and $\sigma \in A_n$ fixes exactly one letter, then $(A_n)_\sigma$ forces congruence permutability. However, Corollary 2.15 cannot be used to prove that $(A_n)_\sigma$ forces congruence permutability when $n = 4, 5, 4k+2$ or $4k+3$ and $\sigma \in A_n$ fixes at most 3 letters. In each of these cases, some nucleus has a retraction onto $\langle\sigma\rangle$.

Example 3.13. The Suzuki groups $\mathrm{Sz}(q)$, $q = 2^{2m+1}$ and $m > 0$, are nonabelian simple groups. Let $\sigma \in \mathrm{Sz}(q)$ be a 2-central involution. We will show that the only nucleus of $G_\sigma := \mathrm{Sz}(q)_\sigma$ is C, and that this nucleus has no retraction onto $\langle\sigma\rangle$. This will prove that $\mathrm{Sz}(q)_\sigma$ forces congruence permutability.

We will need the following Facts about Suzuki groups:

(1) If p is an odd prime, then the p-Sylow subgroups of G are cyclic (Theorem 3.9 of [7]).
(2) If P_2 is a 2-Sylow subgroup of G, then $Z(P_2) = [P_2, P_2]$, and both $P_2/[P_2, P_2]$ and $[P_2, P_2]$ are elementary abelian 2-groups of cardinality q (from Theorem 3.10(c) and Lemma 3.1 of [7]).
(3) If P_2 and $\overline{P}_2$ are distinct 2-Sylow subgroups of G, then $P_2 \cap \overline{P}_2 = \{1\}$ (Theorem 13.9 of [7]).

The property described in Fact (3) is inherited by subgroups, so the 2-Sylow subgroups of C are pairwise disjoint. But they all contain σ, so C contains a unique 2-Sylow subgroup, P_2.

Suppose that the subgroup $H \leq G$ is nontrivial. As noted at the beginning of this section, $HH^\sigma \cong H \times H$. Since HH^σ is a square subgroup, none of its Sylow subgroups are embeddable into cyclic groups. From Fact (1), HH^σ must be a 2-group, hence can be extended to a 2-Sylow subgroup $\overline{P}_2 \leq G$. The Sylow subgroups P_2 and $\overline{P}_2$ must be different, since $\sigma \in Z(P_2)$ but σ does not centralize $H \subseteq \overline{P}_2$. Now, the 2-group $C \cap (HH^\sigma)$ is contained in $\overline{P}_2$, since HH^σ is, and it is contained in P_2, since $C \cap (HH^\sigma)$ is a 2-group that is contained in C and P_2 is the unique 2-Sylow subgroup of C. Thus, by Fact (3), $C \cap (HH^\sigma) \subseteq P_2 \cap \overline{P}_2 = \{1\}$. Since $\rho\colon C \to C$ is a retraction onto $C \cap (HH^\sigma)$, ρ is constant, hence $K = \ker(\rho) = C$.

Now suppose that ρ' is a retraction of $K = C$ onto $\langle\sigma\rangle$. Since $\langle\sigma\rangle \leq P_2 \leq C = K$, ρ' restricts to a retraction of P_2 onto $\langle\sigma\rangle$. Necessarily $\ker(\rho'|_{P_2})$ is a subgroup of index 2 in P_2 which does not contain σ. But any subgroup of index 2 in P_2 contains $[P_2, P_2] = Z(P_2)$, and $\sigma \in Z(P_2)$. Thus K has no retraction onto $\langle\sigma\rangle$. Corollary 2.15 applies to show that $\mathrm{Sz}(q)_\sigma$ forces congruence permutability.

Example 3.14. Here we show that if σ is a 2-central involution in the Mathieu group M_{11}, then $(M_{11})_\sigma$ forces congruence permutability.

The only fact that we will need about this group is that if σ is a 2-central involution, then $C \cong \mathrm{GL}(2, 3)$ (Theorem 5.2 of [7]). Since $-I$ is the unique central involution of $\mathrm{GL}(2, 3)$, σ must correspond to $-I$ under this isomorphism. The retraction ρ'

of $\mathrm{GL}(2,3)$ that corresponds to ρ must restrict to a retraction of the commutator subgroup $\mathrm{SL}(2,3)$ into itself satisfying $\rho'(-I) = I$. Hence $-I \notin \mathrm{im}(\rho')$. But $-I$ is the unique involution of $\mathrm{SL}(2,3)$, forcing $\rho'(\mathrm{SL}(2,3))$ to have odd order. This proves that $\ker(\rho')$ contains the unique 2-Sylow subgroup of $\mathrm{SL}(2,3)$, which is an 8-element quaternion group containing $-I$ in its center. Applying this information to ρ, we conclude that $K = \ker(\rho)$ contains an 8-element quaternion group with σ in its center. The fact that the quaternion group has no retraction onto its center prevents K from having a retraction onto $\langle\sigma\rangle$. Thus $(M_{11})_\sigma$ forces congruence permutability.

Remark 3.15. Using GAP, we have checked that when σ is a 2-central involution of any of the other simple Mathieu groups, M_i, $i = 12, 22, 23, 24$, then no nucleus of M_i retracts onto $\langle\sigma\rangle$. Hence each $(M_i)_\sigma$ forces congruence permutability when σ is 2-central.

4. Class size dimension

It follows from the equations in Definition 2.1 that for any $\mathbf{C} \in \mathcal{V}[G_\sigma]$ there is an action of G on the set $C \times C$ in which σ acts by switching coordinates, namely the action defined by the rule

$$\alpha(a,b) := (t_\alpha(a,b), t_{\sigma\alpha}(a,b)). \tag{4.1}$$

In this section we develop machinery to help determine when pairs $(a,b),(c,d) \in C \times C$ lie in the same G-orbit.

The results of this paper would be stronger and easier to prove if $\mathcal{V}[G_\sigma]$ was congruence uniform.[1] It isn't, so we proceed to introduce a measure of the nonuniformity of congruences. In the typical situation considered, $\mathbf{B}$ and $\mathbf{C}$ are finite algebras in $\mathcal{V}[G_\sigma]$ and $f\colon \mathbf{B} \to \mathbf{C}$ is a surjective homomorphism. The function $C \to \mathbb{Z}\colon c \mapsto |f^{-1}(c)|$ lists the class sizes of the congruence $\ker(f)$. We will study the real vector space $\mathrm{Vect}(\mathbf{C})$ spanned by the functions of the form

$$\phi_f\colon C \to \mathbb{R}\colon c \mapsto \log(|f^{-1}(c)|)\,, \tag{4.2}$$

as $\mathbf{B}$ ranges over all finite algebras in $\mathcal{V}[G_\sigma]$ and f ranges over all surjective homomorphisms $f\colon \mathbf{B} \to \mathbf{C}$. The vector space dimension of $\mathrm{Vect}(\mathbf{C})$, called the *class size dimension over* $\mathbf{C}$ and denoted $\dim(\mathbf{C})$, is a measure of the nonuniformity of the kernels of homomorphisms onto $\mathbf{C}$.

If $\mathbf{C} \in \mathcal{V}[G_\sigma]$ is a nonempty finite algebra and $\mathbf{D} \in \mathcal{V}[G_\sigma]$ is a nontrivial algebra, then the surjective homomorphism $\pi_1\colon \mathbf{C} \times \mathbf{D} \to \mathbf{C}\colon (c,d) \to c$ induces the nonzero constant function $\phi_{\pi_1}(c) = \log(|D|) \in \mathrm{Vect}(\mathbf{C})$. This shows that $\dim(\mathbf{C}) \geq 1$ for any nonempty finite algebra in $\mathcal{V}[G_\sigma]$. It also shows that $\dim(\mathbf{C}) = 1$ if and only if

[1] A congruence is *uniform* if all classes have the same size. An algebra is congruence uniform if all of its congruences are uniform, and a variety is congruence uniform if all of its algebras are.

$\mathrm{Vect}(\mathbf{C})$ contains only constant functions, which holds if and only if the kernel of any surjective homomorphism $f\colon \mathbf{B} \to \mathbf{C}$, with $\mathbf{B} \in \mathcal{V}[G_\sigma]$ finite, is a uniform congruence.

Lemma 4.1. *Let $\mathbf{C}$ be a finite algebra in $\mathcal{V}[G_\sigma]$. If $\phi \in$ Vect$(\mathbf{C})$, $a, b \in C$, and $\alpha \in G$, then*

$$\phi(a) + \phi(b) = \phi(t_\alpha(a,b)) + \phi(t_{\sigma\alpha}(a,b))\,. \tag{4.3}$$

Proof. Referring back to (4.2), $\mathrm{Vect}(\mathbf{C})$ is spanned by vectors of the form ϕ_f where f is a surjective homomorphism $\mathbf{B} \to \mathbf{C}$. Therefore it is enough to verify (4.3) when $\phi = \phi_f$. In this case, the left hand side of (4.3) is

$$\phi_f(a) + \phi_f(b) = \log(|f^{-1}(a)|) + \log(|f^{-1}(b)|) = \log(|f^{-1}(a) \times f^{-1}(b)|),$$

while the right hand side is

$$\phi_f(t_\alpha(a,b)) + \phi_f(t_{\sigma\alpha}(a,b)) = \log(|f^{-1}(t_\alpha(a,b)) \times f^{-1}(t_{\sigma\alpha}(a,b))|).$$

To establish (4.3) it suffices to show that

$$|f^{-1}(a) \times f^{-1}(b)| = |f^{-1}(t_\alpha(a,b)) \times f^{-1}(t_{\sigma\alpha}(a,b))|.$$

We prove this by showing that the function $(c,d) \mapsto (t_\alpha(c,d), t_{\sigma\alpha}(c,d))$ is a bijection from $f^{-1}(a) \times f^{-1}(b)$ onto $f^{-1}(t_\alpha(a,b)) \times f^{-1}(t_{\sigma\alpha}(a,b))$. Choose $(c,d) \in f^{-1}(a) \times f^{-1}(b)$. Then $f(c) = a$ and $f(d) = b$. Thus $f(t_\alpha(c,d)) = t_\alpha(f(c), f(d)) = t_\alpha(a,b)$, meaning that $t_\alpha(c,d) \in f^{-1}(t_\alpha(a,b))$, and similarly $t_{\sigma\alpha}(c,d) \in f^{-1}(t_{\sigma\alpha}(a,b))$. This proves that, in terms of the G-action on $B \times B$ defined by line (4.1), α maps pairs of $f^{-1}(a) \times f^{-1}(b)$ to pairs of $f^{-1}(t_\alpha(a,b)) \times f^{-1}(t_{\sigma\alpha}(a,b))$. The inverse map is provided by α^{-1}. □

Theorem 4.2. *Let G_σ be a group with designated involution σ. Let $\mathbf{F}$ be a finite, 2-generated, free algebra in $\mathcal{V}[G_\sigma]$ for which the canonical homomorphism $\kappa\colon G_\sigma \to$ Aut$(\mathbf{F})_\sigma$ is an isomorphism. Let $\mathcal{V}$ be the variety generated by $\mathbf{F}$. If $\mathbf{C} \in \mathcal{V}$ is finite and $(a,b),(c,d) \in C \times C$, then (a,b) and (c,d) belong to the same G-orbit, under the action defined in line (4.1), if and only if*

1. *$\{a,b\}$ and $\{c,d\}$ generate the same subalgebra $\mathbf{D}$ of $\mathbf{C}$, and*
2. *$\phi(a) + \phi(b) = \phi(c) + \phi(d)$ for all $\phi \in$ Vect$(\mathbf{D})$.*

Proof. Assume first that (a,b) and (c,d) belong to the same G-orbit. Then there is an $\alpha \in G$ such that $(c,d) = \alpha(a,b) = (t_\alpha(a,b), t_{\sigma\alpha}(a,b))$. This shows that $c = t_\alpha(a,b)$ and $d = t_{\sigma\alpha}(a,b)$, so c and d belong to subalgebra generated by $\{a,b\}$. Similarly, $(a,b) = \alpha^{-1}(c,d)$, so a and b belong to the subalgebra generated by $\{c,d\}$. This proves that (1) must hold. That (2) holds follows from Lemma 4.1.

Now we assume that (1) and (2) hold and argue that $(c,d) = (t_\alpha(a,b), t_{\sigma\alpha}(a,b))$ for some α. By (1), the homomorphism $f\colon \mathbf{F} \to \mathbf{D}$ that is defined on the generators $x, y \in F$ by $x \mapsto a$ and $y \mapsto b$ is surjective. For each subuniverse $S \leq F$ let

$$N(S; r, s) = |(f^{-1}(r) \cap S) \times (f^{-1}(s) \cap S)|$$

denote the number of pairs $(p, q) \in S \times S$ such that $f(p) = r$ and $f(q) = s$.

Claim 4.3. *$N(S; a, b) = N(S; c, d)$ for all subuniverses S of F.*

Proof. If $f|_S \colon \mathbf{S} \to \mathbf{D}$ is not surjective, then since $\{a, b\}$ and $\{c, d\}$ are both generating sets there can be no pairs $(p, q) \in S \times S$ such that $f(p) = a$ and $f(q) = b$, or $f(p) = d$ and $f(q) = d$. That is, $N(S; a, b) = N(S; c, d) = 0$ when $f(S) \neq D$.

Now suppose that $g := f|_S \colon \mathbf{S} \to \mathbf{D}$ is surjective. Then

$$\log(N(S; a, b)) = \log(|g^{-1}(a) \times g^{-1}(b)|) = \phi_g(a) + \phi_g(b),$$

and similarly $\log(N(S; c, d)) = \phi_g(c) + \phi_g(d)$. Hence $N(S; a, b) = N(S; c, d)$ is a consequence of our assumption that (2) holds, □

The pairs (u, v) in $f^{-1}(a) \times f^{-1}(b)$ for which $\{u, v\}$ is a generating set for $\mathbf{F}$ are precisely the pairs in $f^{-1}(a) \times f^{-1}(b)$ that lie in no proper subuniverse $S \subsetneqq F$. Therefore, by the principle of inclusion and exclusion, the number of pairs $(u, v) \in f^{-1}(a) \times f^{-1}(b)$ for which $\{u, v\}$ is a generating set for $\mathbf{F}$ is

$$N(F; a, b) - \sum_{\mathcal{S}' \subseteq \mathcal{S}} (-1)^{|\mathcal{S}'|+1} N\left(\bigcap \mathcal{S}'; a, b\right)$$

where $\mathcal{S}$ is the family of all proper subuniverses of $\mathbf{F}$ and $\mathcal{S}'$ ranges over nonempty subfamilies of $\mathcal{S}$. By Claim 4.3, $N(S; a, b) = N(S; c, d)$ for any subuniverse $S \subseteq F$; therefore the number of pairs $(u, v) \in f^{-1}(a) \times f^{-1}(b)$ such that $\{u, v\}$ is a generating set for $\mathbf{F}$ is the same as the number of pairs $(p, q) \in f^{-1}(c) \times f^{-1}(d)$ such that $\{p, q\}$ is a generating set for $\mathbf{F}$. But there is at least one pair $(u, v) \in f^{-1}(a) \times f^{-1}(b)$ such that $\{u, v\}$ is a generating set, namely $(u, v) = (x, y)$. Therefore, there is at least one pair $(p, q) \in f^{-1}(c) \times f^{-1}(d)$ such that $\{p, q\}$ is a generating set for $\mathbf{F}$.

If $\{p, q\}$ is a generating set for $\mathbf{F}$, then the endomorphism of $\mathbf{F}$ defined on generators by $x \mapsto p$ and $y \mapsto q$ is surjective. Since $\mathbf{F}$ is finite, this endomorphism is an automorphism. Since the canonical homomorphism κ is an isomorphism, this endomorphism is $\kappa(\alpha)$ for some $\alpha \in G$. But then $(p, q) = (x, y)\kappa(\alpha) = (t_\alpha(x, y), t_{\sigma\alpha}(x, y))$, so $c = f(p) = f(t_\alpha(x, y)) = t_\alpha(f(x), f(y)) = t_\alpha(a, b)$, and similarly $d = t_{\sigma\alpha}(a, b)$. This proves that $(c, d) = \alpha(a, b)$, hence (a, b) and (c, d) lie in the same G-orbit. □

If $d = \dim(\mathbf{C})$, then there is a basis $\mathcal{B} = \{\phi_0, \ldots, \phi_{d-1}\}$ for $\mathrm{Vect}(\mathbf{C})$ with ϕ_0 equal to the nonzero constant function with range $\{1\}$. The function

$$\Phi \colon C \to \mathbb{R}^d \colon c \mapsto (\phi_0(c), \ldots, \phi_{d-1}(c)) \tag{4.4}$$

maps the finite set C into $\mathbb{R}^d$. The convex hull of the finite set of points $\Phi(C)$ is a polytope in $\mathbb{R}^d$, which we will denote $\mathrm{Poly}(\mathbf{C})$. This polytope depends on the choice of the basis $\mathcal{B}$ for $\mathrm{Vect}(\mathbf{C})$, but we will see that some of its properties — such as its dimension — do not.

To fix language, a *hyperplane* in $\mathbb{R}^d$ is a set of the form

$$H = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{u}^t\mathbf{x} = r\} \tag{4.5}$$

for some fixed $\mathbf{u} \in \mathbb{R}^d$ and some fixed $r \in \mathbb{R}$. The set $\{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{u}^t\mathbf{x} \leq r\}$ is a *half space* determined by the hyperplane (4.5). A *polytope* in $\mathbb{R}^d$ may be equivalently defined as either the convex hull of a finite set of points or as a bounded set that is the intersection of finitely many half spaces. A hyperplane H *supports* a polytope P if P lies in a half space determined by H. A *face* of a polytope P is the intersection of P with one of its supporting hyperplanes. A face may be empty. The *face lattice* of a polytope is the lattice of all faces ordered by inclusion. It can be shown that the intersection of two faces is a face, so the meet in this lattice is intersection. The *dimension* of a (face of a) polytope is the least dimension of an affine subspace containing the (face of the) polytope. A face of dimension zero is a *vertex*, and a face of dimension one is an *edge*.

To compute the dimension of $\mathrm{Poly}(\mathbf{C})$, note that each point $\Phi(c)$ lies in the affine subspace of $\mathbb{R}^d$ consisting of tuples with first coordinate 1 (since $\phi_0 \equiv 1$). Thus, the dimension of $\mathrm{Poly}(\mathbf{C})$ is at most $d-1$. If the dimension of $\mathrm{Poly}(\mathbf{C})$ is $< d-1$, then $\mathrm{Poly}(\mathbf{C})$ must lie in a proper linear subspace of $\mathbb{R}^d$, hence there is a nonzero vector $(A_0, \ldots, A_{d-1})^t \in \mathbb{R}^d$ orthogonal to all tuples $\Phi(c)$, $c \in C$. This implies that $A_0\phi_0 + \cdots + A_{d-1}\phi_{d-1} = 0$ is a dependence relation among the elements of $\mathcal{B}$, contradicting the fact that $\mathcal{B}$ is a basis. Therefore, the dimension of $\mathrm{Poly}(\mathbf{C})$ is exactly $d-1 = \dim(\mathbf{C})-1$, which allows us to use $\mathrm{Poly}(\mathbf{C})$ to compute the class size dimension over $\mathbf{C}$.

Lemma 4.4. *Let G be a group with a designated involution σ, and let $\mathbf{C}$ be a finite algebra in $\mathcal{V}[G_\sigma]$. If F is a face of Poly($\mathbf{C}$), then $\Phi^{-1}(F)$ is a subuniverse of $\mathbf{C}$. In fact, the function Φ^{-1} is a meet embedding of face lattice of Poly($\mathbf{C}$) into the subalgebra lattice $\mathbf{Sub}(\mathbf{C})$.*

Proof. Let F be a face of $\mathrm{Poly}(\mathbf{C})$. Suppose that $a, b \in \Phi^{-1}(F)$, or equivalently that $\Phi(a), \Phi(b) \in F$. Choose any $\alpha \in G$. It follows from Lemma 4.1 that

$$\Phi(a) + \Phi(b) = \Phi(t_\alpha(a,b)) + \Phi(t_{\sigma\alpha}(a,b))\,. \tag{4.6}$$

Since $(\Phi(a) + \Phi(b))/2$ is the midpoint of the segment joining two elements of F, it belongs to F. Equation (4.6) implies that the midpoint of the segment joining $\Phi(t_\alpha(a,b))$ and $\Phi(t_{\sigma\alpha}(a,b))$ therefore lies in F, so each of these points lie in F since F is a face. In particular, $t_\alpha(a,b) \in \Phi^{-1}(F)$. Since a, b and α were arbitrary, $\Phi^{-1}(F)$ is a subuniverse.

Since the meet operation in both the face lattice of $\mathrm{Poly}(\mathbf{C})$ and in $\mathbf{Sub}(\mathbf{C})$ is intersection, it follows that Φ^{-1} preserves meet.

To see that Φ^{-1} is 1-1, assume that $F \not\subseteq F'$ are faces. Then some vertex of F does not belong to F'. But a vertex of a face is a vertex of $\mathrm{Poly}(\mathbf{C})$, and all of these are in the image of Φ. Thus $\Phi^{-1}(F) \not\subseteq \Phi^{-1}(F')$. □

Corollary 4.5. *If k is the largest number of nonempty subuniverses in any chain $\emptyset \neq U_1 \subsetneqq U_2 \subsetneqq \cdots \subsetneqq U_k$ in $\mathbf{Sub}(\mathbf{C})$, then $\dim(\mathbf{C}) \leq k$.*

Proof. If $\dim(\mathbf{C}) = d$, then the dimension of $\mathrm{Poly}(\mathbf{C})$ is $d-1$. A $(d-1)$-dimensional polytope has faces $\emptyset \neq F_1 \subsetneqq \cdots \subsetneqq F_d$ where the dimension of F_i is $i-1$. By Lemma 4.4, $\emptyset \neq \Phi^{-1}(F_1) \subsetneqq \cdots \subsetneqq \Phi^{-1}(F_d)$ is a chain of d nonempty subuniverses of $\mathbf{C}$. Since k is the length of the longest chain of subuniverses, $\dim(\mathbf{C}) = d \leq k$. □

Lemma 4.6. *Let G be a group with a designated involution σ, let $\mathbf{C}$ be a finite algebra in $\mathcal{V}[G_\sigma]$, and let $\mathbf{S} \leq \mathbf{C}$ be a nonempty subalgebra. If $\phi \in$ Vect($\mathbf{C}$), then $\phi|_S \in$ Vect($\mathbf{S}$).*

In particular, if the function Φ defined in line (4.4) is nonconstant on S, then $\dim(\mathbf{S}) \geq 2$.

Proof. To prove that $\phi \in \mathrm{Vect}(\mathbf{C})$ implies $\phi|_S \in \mathrm{Vect}(\mathbf{S})$, it suffices to prove it on a basis for $\mathrm{Vect}(\mathbf{C})$. Hence we may assume that $\phi = \phi_f$ for some surjective homomorphism f from some finite $\mathbf{B} \in \mathcal{V}[G_\sigma]$ to $\mathbf{C}$.

Let $\mathbf{B}' = f^{-1}(S)$ and let $g = f|_{B'}$. Then g is a surjective homomorphism from some finite $\mathbf{B}' \in \mathcal{V}[G_\sigma]$ to $\mathbf{S}$, so $\phi_g \in \mathrm{Vect}(\mathbf{S})$. But $\phi_g = \phi_f|_S$.

The second claim of the lemma follows from the first and the fact that $\mathrm{Vect}(\mathbf{S})$ contains a nonzero constant vector whenever $\mathbf{S}$ is nonempty. □

Lemma 4.7. *Let G be a group with a designated involution σ and let $\mathbf{C} \in \mathcal{V}[G_\sigma]$ be finite. If $\gamma \in$ Aut($\mathbf{C}$) and $\phi \in$ Vect($\mathbf{C}$), then $\phi \circ \gamma^{-1} \in$ Vect($\mathbf{C}$).*

Proof. Suppose that $\mathbf{B} \in \mathcal{V}[G_\sigma]$ is finite, $f\colon \mathbf{B} \to \mathbf{C}$ is surjective, and $\phi = \phi_f$. The composite homomorphism $g\colon \mathbf{B} \xrightarrow{f} \mathbf{C} \xrightarrow{\gamma} \mathbf{C}$ induces the vector $\phi_g = \phi_f \circ \gamma^{-1}$. This shows that the implication

$$\phi \in \mathrm{Vect}(\mathbf{C}) \implies \phi \circ \gamma^{-1} \in \mathrm{Vect}(\mathbf{C})$$

holds when $\phi = \phi_f$, hence it holds for all $\phi \in \mathrm{Vect}(\mathbf{C})$. □

Now we consider the class size dimensions over algebras $\mathbf{C}$ from a minimal, locally finite, congruence permutable variety $\mathcal{M} \leq \mathcal{V}[G_\sigma]$. Recall the structure of algebras in such varieties: an algebra is *strictly simple* if it is finite, simple, and has no nontrivial proper subuniverses. A minimal locally finite variety has the form $\mathcal{M} = \mathsf{HSP}(\mathbf{A})$ for some uniquely determined strictly simple algebra $\mathbf{A}$ (cf. [8]). If $\mathcal{M}$ is congruence permutable, then $\mathbf{A}$ is either a quasiprimal algebra or an affine algebra with a 1-element subuniverse (cf. Theorems 12.1 and 12.4 of [2]). Whether $\mathbf{A}$ is quasiprimal or affine, Fleischer's Lemma guarantees that each finite algebra $\mathbf{C} \in \mathcal{M}$ is isomorphic to $\mathbf{A}^n$ for some finite n.

Lemma 4.8. *Let* $\mathbf{A}$ *be a strictly simple quasiprimal algebra. If* S *is a maximal proper subuniverse of* $\mathbf{A}^n$ *for some* $n \geq 1$*, then either* $S = A \times \cdots \times \underbrace{\{0\}}_{i} \times \cdots \times A$ *for some* i *and some* 1*-element subuniverse* $\{0\}$*, or*

$$S = \{(a_1, \ldots, a_n) \mid a_i = a_j\alpha \text{ for some } i \neq j \text{ and some } \alpha \in \mathrm{Aut}(\mathbf{A})\}. \tag{4.7}$$

Proof. This follows from Theorem 4.2 of [9], which characterizes the subuniverses of finite powers of quasiprimal algebras. □

Theorem 4.9. *Let* $\mathcal{M} \leq \mathcal{V}[G_\sigma]$ *be a minimal, locally finite, congruence permutable variety with strictly simple generator* $\mathbf{A}$*. If* $\dim(\mathbf{A}) = 1$*, then* $\dim(\mathbf{A}^n) = 1$ *for all* $n \geq 1$*.*

Proof. Assume not. Let $n > 1$ be the least positive integer such that $\dim(\mathbf{A}^n) > 1$. The first fact to establish is that $\dim(\mathbf{A}^n) = 2$.

If instead $\dim(\mathbf{A}^n) > 2$, then $\mathrm{Poly}(\mathbf{A}^n)$ has dimension greater then 1, so there is a proper face $F \subsetneqq \mathrm{Poly}(\mathbf{A}^n)$ of dimension exactly 1 (an edge). According to Lemma 4.4, $S := \Phi^{-1}(F)$ is a proper subuniverse of $\mathbf{A}^n$. $\Phi(S)$ contains the distinct endpoints of F, so Φ is nonconstant on S. By Lemma 4.6, $\dim(\mathbf{S}) \geq 2$. But $\mathbf{S} \cong \mathbf{A}^k$ for some $k < n$, contradicting the minimality of n. The conclusion is that $\dim(\mathbf{A}^n) = 2$ and $\dim(\mathbf{S}) = 1$ for all proper subalgebras $\mathbf{S} \leq \mathbf{A}^n$.

$\mathrm{Poly}(\mathbf{A}^n)$ has dimension one, so it is an edge. Let u and v be the endpoints, and let $U = \Phi^{-1}(\{u\})$ and $V = \Phi^{-1}(\{v\})$ be the proper nonempty subuniverses of $\mathbf{A}^n$ that map to u and v respectively. These subuniverses are disjoint since $U \cap V = \Phi^{-1}(\{u\} \cap \{v\}) = \Phi^{-1}(\emptyset) = \emptyset$. They are also maximal subuniverses, since if U' is a subuniverse of $\mathbf{A}^n$ that properly extends U, then Φ is nonconstant when restricted to U'. By Lemma 4.6, $\dim(\mathbf{U}') \geq 2$, so $U' = A^n$. The maximal subuniverses of $\mathbf{A}^n$, when $\mathbf{A}$ is a strictly simple algebra in a congruence permutable variety, are isomorphic to $\mathbf{A}^{n-1}$.

Suppose that $\mathbf{A}$ has a 1-element subuniverse $\{0\}$. If U contains no tuple $(a_1, \ldots, a_n)$ with $a_i = 0$ for some i, then the restrictions of the projections $\pi_i \colon \mathbf{A}^n \to \mathbf{A}$ to U fail to be surjective for every i. Since $\mathbf{A}$ is strictly simple and $\pi_i(U)$ is a proper subuniverse of $\mathbf{A}$, the restriction of each π_i to U is constant, forcing $|U| = 1$. This contradicts $|U| = |A|^{n-1}$. Hence U contains a tuple $(a_1, \ldots, a_n)$ with $a_i = 0$ for some i, and similarly V contains a tuple $(b_1, \ldots, b_n)$ with $b_j = 0$ for some j. Let W be the subuniverse of $\mathbf{A}^n$ of all tuples $(a_1, \ldots, a_n)$ with $a_i = 0$, and let X be the subuniverse of $\mathbf{A}^n$ of all tuples $(b_1, \ldots, b_n)$ with $b_j = 0$. Then, since U, V, W and X are proper subuniverses, Φ is constant on each of these subuniverses. But U intersects W, which intersects X, which intersects V, so Φ is constant on their union. This contradicts the fact that $\Phi(U) = \{u\}$ and $\Phi(V) = \{v\}$. The conclusion is that $\mathbf{A}$ has no 1-element subuniverse. In particular, we are not in the case where $\mathbf{A}$ is an affine

algebra with a 1-element subuniverse, so we must be in the case where $\mathbf{A}$ is a strictly simple quasiprimal algebra.

From Lemma 4.8, $U = \{(a_1, \dots, a_n) \mid a_i = a_j\alpha\}$ and $V = \{(a_1, \dots, a_n) \mid a_k = a_\ell\beta\}$. Since U and V are disjoint, $\{i, j\} = \{k, \ell\}$. If $n > 2$, then there is a third coordinate $k \notin \{i, j\}$, and $W = \{(a_1, \dots, a_n) \mid a_i = a_k\}$ intersects both U and V. This means that Φ is constant on U, V and W, hence on their union. But Φ is not constant on $U \cup V$ since $\Phi(U \cup V) = \{u, v\}$, a contradiction. The conclusion is that $n = 2$.

Since $\dim(\mathbf{A}^2) = 2$, the function $\Phi(x)$ is of the form $(\phi_0(x), \phi_1(x))$ where ϕ_0 is the constant vector with range $\{1\}$. There is no harm in assuming that $\phi_1 = \phi_f$ is a nonconstant vector in $\mathrm{Vect}(\mathbf{A}^2)$ induced by some surjective homomorphism $f\colon \mathbf{B} \to \mathbf{A}^2$ from some finite $\mathbf{B} \in \mathcal{V}[G_\sigma]$, so we make that assumption. This meaning for f and ϕ_f should now be considered fixed for the remainder of the proof. The endpoints u and v of $\mathrm{Poly}(\mathbf{A}^2)$ have the form $u = (1, r)$ and $v = (1, s)$ where r and s are the least and largest real numbers in the set $\phi_f(A^2)$ in some order. The function ϕ_f defines a linear quasiorder on A^2 by $(a, b) \leq (c, d)$ if and only if $\phi_f(a, b) \leq \phi_f(c, d)$ in $\mathbb{R}$. The subuniverses $U = \Phi^{-1}(\{u\}) = \phi_f^{-1}(r)$ and $V = \Phi^{-1}(\{v\}) = \phi_f^{-1}(s)$ are the sets of least and largest elements in A^2 under this quasiorder. In particular,

$$\phi_f(U) = \{r\} \neq \{s\} = \phi_f(V). \tag{4.8}$$

Select another nonconstant vector $\phi \in \mathrm{Vect}(\mathbf{A}^2)$, and use it to define a second linear quasiorder on A^2: $(a, b) \ll (c, d)$ if and only if $\phi(a, b) \leq \phi(c, d)$ in $\mathbb{R}$. Since $\phi = c_0\phi_0 + c_f\phi_f$ with $c_f \neq 0$ when ϕ is nonconstant, and since ϕ_0 is a constant vector, it follows that $\ll = \leq$ when $c_f > 0$ and $\ll = \leq^\partial$ when $c_f < 0$. Thus, the linear quasiorder on A^2 arising from a nonconstant vector from $\mathrm{Vect}(\mathbf{A}^2)$ is uniquely determined up to duality.

By Lemma 4.7, if $\Gamma \in \mathrm{Aut}(\mathbf{A}^2)$, then $\phi \circ \Gamma^{-1} \in \mathrm{Vect}(\mathbf{A}^2)$. The vector $\phi \circ \Gamma^{-1}$ is nonconstant since ϕ is nonconstant and Γ is a permutation. Therefore the linear quasiorder defined by $\phi \circ \Gamma^{-1}$ is either $\leq$ or $\leq^\partial$. If it is $\leq$, then Γ preserves $\leq$, while if it is $\leq^\partial$, then Γ reverses $\leq$. Since the sets of least and largest elements under the quasiorder are U and V (in some order), this forces either $(U)\Gamma = U$ and $(V)\Gamma = V$, or else $(U)\Gamma = V$ and $(V)\Gamma = U$.

Both U and V are maximal proper subuniverses of $\mathbf{A}^2$, and $\mathbf{A}$ is a quasiprimal algebra without proper subuniverses, so from Lemma 4.8 it follows that U and V are graphs of automorphisms of $\mathbf{A}$. Say $U = \{(a, a\alpha) \mid \alpha \in \mathrm{Aut}(\mathbf{A})\}$ and $V = \{(a, a\beta) \mid \beta \in \mathrm{Aut}(\mathbf{A})\}$. If $\gamma \in \mathrm{Aut}(\mathbf{A})$, then $\Gamma = id \times \gamma \in \mathrm{Aut}(\mathbf{A}^2)$. By the result of the previous paragraph, Γ maps each of U and V into itself or else it interchanges the two. Since $(U)\Gamma = \{(a, a(\alpha\gamma)) \mid a \in A\}$ is the graph of $\alpha\gamma$ and $(V)\Gamma$ is the graph of $\beta\gamma$, this implies that $\{\alpha\gamma, \beta\gamma\} = \{\alpha, \beta\}$. Since this holds for every $\gamma \in \mathrm{Aut}(\mathbf{A})$, it must be that $\mathrm{Aut}(\mathbf{A}) = \{\alpha, \beta\} = \{id, \tau\}$ for some involution τ. Since U and V are the graphs of distinct automorphisms of $\mathbf{A}$, and $|\mathrm{Aut}(\mathbf{A})| = 2$, U and V are the graphs of id and τ.

Let us define some vectors in $\mathrm{Vect}(\mathbf{A}^3)$. As has been our convention, ϕ_0 will denote the constant vector with range $\{1\}$. From above we have a surjective homomorphism $f\colon \mathbf{B} \to \mathbf{A}^2$, which we write here in terms of its components: $f(x) = (f_1(x), f_2(x))$, where $f_i = \pi_i \circ f$. The function $g_1 := id \times f\colon \mathbf{A} \times \mathbf{B} \to \mathbf{A} \times \mathbf{A}^2 = \mathbf{A}^3$ is a surjective homomorphism from an algebra $\mathbf{A} \times \mathbf{B} \in \mathcal{V}[G_\sigma]$ to $\mathbf{A}^3$. In terms of its components, it is the function $g_1(a, b) = (a, f_1(b), f_2(b))$. By cyclically permuting coordinates we obtain other surjective homomorphisms $g_2(a, b) = (f_2(b), a, f_1(b))$ and $g_3(a, b) = (f_1(b), f_2(b), a)$ from $\mathbf{A} \times \mathbf{B}$ to $\mathbf{A}^3$. Let $\phi_{g,1}, \phi_{g,2}$ and $\phi_{g,3}$ be the induced vectors in $\mathrm{Vect}(\mathbf{A}^3)$. Observe that if $\mathbf{a} = (a_1, a_2, a_3) \in A^3$, then

$$\phi_{g,1}(\mathbf{a}) = \log(|g_1^{-1}(\mathbf{a})|) = \log(|\{a_1\} \times f^{-1}(a_2, a_3)|) = \phi_f(a_2, a_3), \tag{4.9}$$

and similarly $\phi_{g,2}(\mathbf{a}) = \phi_f(a_3, a_1)$ and $\phi_{g,3}(\mathbf{a}) = \phi_f(a_1, a_2)$.

The fact that $\mathbf{A}$ is quasiprimal with no 1-element subuniverse implies that the subuniverses of $\mathbf{A}^3$ have size $|A|, |A|^2$ or $|A|^3$. In particular, a maximal chain of nonempty subuniverses of $\mathbf{A}^3$ contains at most 3 members. By Corollary 4.5, $\dim(\mathbf{A}^3) \leq 3$. This means that the set $\{\phi_0, \phi_{g,1}, \phi_{g,2}, \phi_{g,3}\}$ is not linearly independent. Suppose that

$$c_0\phi_0 + c_1\phi_{g,1} + c_2\phi_{g,2} + c_3\phi_{g,3} = 0 \tag{4.10}$$

is a nontrivial dependence relation. It cannot be that $c_1 = c_2 = c_3 = 0$, since $\phi_0 \neq 0$, so we may assume by symmetry that $c_1 \neq 0$. Choose and fix an element $0 \in A$. By applying $\phi_{g,1}$ to a triple of the form $(0, x, y) \in A^3$ we derive from (4.9) that $\phi_{g,1}(0, x, y) = \phi_f(x, y)$. Similarly, $\phi_{g,2}(0, x, y) = \phi_f(y, 0)$ and $\phi_{g,3}(0, x, y) = \phi_f(0, x)$. Thus, if we solve for $\phi_{g,1}$ in equation (4.10), evaluate all functions at $(0, x, y)$, and write everything in terms of ϕ_f we get that

$$\phi_f(x, y) = -(c_2/c_1)\phi_f(y, 0) - ((c_3/c_1)\phi_f(0, x) + (c_0/c_1)) \tag{4.11}$$

proving that $\phi_f(x, y)$ is the sum of two real valued unary functions. For simplicity of notation, we will rewrite (4.11) as $\phi_f(x, y) = h(x) + k(y)$.

By (4.8), $\phi_f(U) = \{r\}$, $\phi_f(V) = \{s\}$, and $r \neq s$. We have established that U and V are the graphs of the automorphisms $id, \tau \in \mathrm{Aut}(\mathbf{A}^2)$. By interchanging U with V and r with s if necessary we may assume that U is the graph of id and V is the graph of τ. Since $\phi_f(U) = \{r\}$, and $U = \{(x, x) \mid x \in A\}$, we get that $\phi_f(x, x) = h(x) + k(x) = r$ for all $x \in A$. Thus, $k(x) = r - h(x)$, and so $\phi_f(x, y) = h(x) - h(y) + r$. The fact that $\phi_f(V) = \{s\}$ and $V = \{(x, x\tau) \mid x \in A\}$ implies that

$$h(x) - h(x\tau) + r = \phi_f(x, x\tau) = s \tag{4.12}$$

for all $x \in A$. Averaging the left and right sides over all $x \in A$, and using that τ is a permutation, yields $r = s$. This contradicts (4.8), and completes the proof. □

If $\mathbf{A}$ is strictly simple, then its nonempty subuniverses can have size 1 or $|A|$ only. Hence $\dim(\mathbf{A}) \leq 2$, according to Corollary 4.5. Theorem 4.9 gives all the information we will need in the case where $\dim(\mathbf{A}) = 1$, so we turn now to the case $\dim(\mathbf{A}) = 2$.

Theorem 4.10. *Let $\mathcal{M} \leq \mathcal{V}[G_\sigma]$ be a minimal, locally finite, congruence permutable variety with strictly simple generator $\mathbf{A}$. If $\dim(\mathbf{A}) = 2$, then $\dim(\mathbf{A}^n) = n+1$ for all $n \geq 1$. Moreover, $\mathbf{A}$ has a linear quasiorder $\leq$ such that*

1. *$\mathbf{A}$ has exactly one minimal element 0 under $\leq$, and $\{0\}$ is a subuniverse. $\mathbf{A}$ has exactly one maximal element 1 under $\leq$, and $\{1\}$ is a subuniverse.*
2. *If $P \subseteq A$ is the set of covers of 0, then $(\{0\} \times P) \cup (P \times \{0\}) \subseteq A \times A$ is a union of orbits under the action defined in line (4.1).*
3. *$\mathbf{A}$ has at most one non-identity automorphism. If it exists, then it is order-reversing with respect to $\leq$.*

Proof. We discuss the linear quasiorder first. Since $\dim(\mathbf{A}) = 2$, $\Phi(x)$ may be taken to be of the form $(\phi_0(x), \phi_f(x))$ where ϕ_0 is the constant vector with range $\{1\}$ and ϕ_f is induced by some surjective homomorphism $f\colon \mathbf{B} \to \mathbf{A}$ from some finite $\mathbf{B} \in \mathcal{V}[G_\sigma]$. Necessarily $\mathrm{Poly}(\mathbf{A})$ is an edge. If its endpoints are $u = (1, r) \in \mathbb{R}^2$ and $v = (1, s) \in \mathbb{R}^2$. then r and s are the least and largest real numbers in range of ϕ_f. The vector ϕ_f defines a linear quasiorder on A by $a \leq b$ if and only if $\phi_f(a) \leq \phi_f(b)$ in $\mathbb{R}$. This is the quasiorder referred to in the statement of the theorem. $U = \Phi^{-1}(\{u\}) = \phi_f^{-1}(r)$ and $V = \Phi^{-1}(\{v\}) = \phi_f^{-1}(s)$ are the sets of least and largest elements in A under this quasiorder. Since endpoints are faces, U and V are proper subuniverses of $\mathbf{A}$, hence $U = \{0\}$ and $V = \{1\}$ for some $0, 1 \in A$. Here the notation is chosen so that $\phi_f(0) < \phi_f(1)$, equivalently $0 < 1$ in the linear quasiorder on A. This establishes property (1) of the linear quasiorder.

Let $P \subseteq A$ be the set of covers of 0, i.e. $p \in P$ if and only if $t := \phi_f(p)$ is the second smallest value attained by ϕ_f. (Recall that $\phi_f(0) = r$ is the smallest value attained by ϕ_f.) If $(p, q) \in X = (\{0\} \times P) \cup (P \times \{0\})$, then $\phi_f(p) + \phi_f(q) = r + t$. If $(m, n) = \alpha(p, q) = (t_\alpha(p, q), t_{\sigma\alpha}(p, q)))$, then from Lemma 4.1 we get that $\phi_f(m) + \phi_f(n) = r + t$. But this forces one of the elements m or n to be 0 and the other to be from P. Thus, if $(p, q) \in X$, then $\alpha(p, q) \in X$ for any $\alpha \in G$.

As argued in the seventh and eighth paragraphs of the proof of Theorem 4.9, the linear quasiorder is unique up to duality and any automorphism of $\mathbf{A}$ either preserves the quasiorder or reverses it. An automorphism that preserves the linear quasiorder must fix the least and largest elements 0 and 1. Since the set of fixed points of an automorphism is a subuniverse, and $\mathbf{A}$ is strictly simple, only the identity automorphism can preserve the linear quasiorder. This means that any two automorphisms that reverse the linear quasiorder are inverses to one another, so there exists at most one automorphism that reverses the linear quasiorder. This establishes (3).

We now argue that $\dim(\mathbf{A}^n) = n+1$ for all finite n. Since a subuniverse of $\mathbf{A}^n$ has cardinality $|A|^k$ for some $0 \leq k \leq n$, Corollary 4.5 proves that $\dim(\mathbf{A}^n) \leq n+1$. We establish equality by exhibiting a set of $n+1$ independent vectors in $\mathrm{Vect}(\mathbf{A}^n)$. The pullback of the homomorphism $f\colon \mathbf{B} \to \mathbf{A}$ along the i-th projection $\pi_i\colon \mathbf{A}^n \to \mathbf{A}$ is

a surjective homomorphism

$$f_i\colon \mathbf{A}\times\cdots\times\underbrace{\mathbf{B}}_{i}\times\cdots\times\mathbf{A}\to\mathbf{A}^n\colon (a_1,\dots,b_i,\dots,a_n)\mapsto(a_1,\dots,f(b_i),\dots,a_n).$$

This function induces the vector $\phi_{f,i}\in\mathrm{Vect}(\mathbf{A}^n)$ defined by $\phi_{f,i}(a_1,\dots,a_n)=\phi_f(a_i)$. Let ϕ_0 be the constant vector with range $\{1\}$. Each one of the vectors in the set $\{\phi_0,\phi_{f,1},\dots,\phi_{f,n}\}$, except $\phi_{f,i}$, is independent of its i-th variable when considered as a function of n variables. Thus, for all i, $\phi_{f,i}$ is not a combination of the remaining vectors. Since ϕ_0 is not the zero vector, the set $\{\phi_0,\phi_{f,1},\dots,\phi_{f,n}\}$ is an independent set of $n+1$ vectors in $\mathrm{Vect}(\mathbf{A}^n)$. Thus, $\dim(\mathbf{A}^n)=n+1$ for all n, as claimed. □

We record some useful information that emerged from the previous proof.

Corollary 4.11. *Let $\mathcal{M}\leq\mathcal{V}[G_\sigma]$ be a minimal, locally finite, congruence permutable variety with strictly simple generator $\mathbf{A}$. Assume that $\dim(\mathbf{A})=2$. If $\phi\in Vect(\mathbf{A})$ is nonconstant, $\phi_0\in Vect(\mathbf{A}^n)$ is the constant vector with range $\{1\}$, and $\phi_i\in Vect(\mathbf{A}^n)$ is defined by*

$$\phi_i(a_1,\dots,a_n):=\phi(a_i),$$

then $\{\phi_0,\phi_1,\dots,\phi_n\}$ is a basis for $Vect(\mathbf{A}^n)$.

Proof. In the proof of Theorem 4.10 this is established in the case where $\phi=\phi_f$. It can be derived for every other nonconstant $\phi\in\mathrm{Vect}(\mathbf{A})$ from the fact that $\phi=c\cdot\phi_0+d\cdot\phi_f$ for some $c,d\in\mathbb{R}$ where $d\neq 0$. □

The next two theorems add further information to Theorem 4.10.

Theorem 4.12. *Let $\mathcal{M}\leq\mathcal{V}[G_\sigma]$ be a minimal, locally finite, congruence permutable variety with strictly simple generator $\mathbf{A}$. If $\dim(\mathbf{A})=2$, then $\mathbf{A}$ is quasiprimal.*

Proof. To prove that $\mathbf{A}$ is quasiprimal, it suffices to show that it is not an affine algebra with a 1-element subuniverse. Assume instead that it is, and that $\{0\}$ is a 1-element subuniverse of $\mathbf{A}$. Then $\mathbf{A}$ is polynomially equivalent to a module with zero element 0. If $p(x,y,z)$ is a Maltsev term for $\mathbf{A}$, then the polynomial $x-y:=p(x,y,0)$ is the subtraction of this module, and

$$\Gamma\colon\mathbf{A}^2\to\mathbf{A}^2\colon(x,y)\mapsto(x-y,y) \tag{4.13}$$

is an automorphism of $\mathbf{A}^2$.

Choose and fix a nonconstant vector $\phi\in\mathrm{Vect}(\mathbf{A})$. From Theorem 4.10 we have $\dim(\mathbf{A}^2)=3$, so according to Corollary 4.11 an explicit basis for $\mathrm{Vect}(\mathbf{A}^2)$ is $\mathcal{B}:=\{\phi_0,\phi_1,\phi_2\}$ where ϕ_0 is the constant vector with range $\{1\}$, and $\phi_i(a_1,a_2)=\phi(a_i)$ for $i=1,2$. Applying Lemma 4.7 to the automorphism Γ defined on line (4.13), we get that $\phi_1\circ\Gamma^{-1}\in\mathrm{Vect}(\mathbf{A}^2)$. Here

$$\phi_1\circ\Gamma^{-1}(x,y)=\phi_1(x+y,y)=\phi(x+y).$$

If we express $\phi_1 \circ \Gamma^{-1}$ in terms of the basis $\mathcal{B}$, we get

$$\phi_1 \circ \Gamma^{-1} = c_0\phi_0 + c_1\phi_1 + c_2\phi_2$$

for certain $c_i \in \mathbb{R}$. Applying both sides to a typical pair $(x, y) \in A^2$, we find that

$$\phi(x + y) = c_0 + c_1\phi(x) + c_2\phi(y).$$

This proves that $\phi(x + y)$ is the sum of a real valued function of $x \in A$ and a real valued function of $y \in A$. Equivalently,

$$\phi(x + y) - \phi(x + 0) - \phi(0 + y) + \phi(0 + 0) = 0,$$

which may be rewritten as

$$\phi(x + y) - \phi(0) = (\phi(x) - \phi(0)) + (\phi(y) - \phi(0)).$$

Thus, $\phi(x) - \phi(0)$ is a homomorphism from the additive group of $\mathbf{A}$ to the additive group of real numbers. Since $\mathbf{A}$ is finite and $\mathbb{R}$ has characteristic zero, the only such function is constant. This is impossible because ϕ was chosen to be nonconstant. This contradiction concludes the proof that $\mathbf{A}$ is not affine. □

The homomorphism μ of the following theorem plays a vital role in Section 6.

Theorem 4.13. *Let G_σ be a group with designated involution σ. Let $\mathbf{F}$ be a finite, 2-generated, free algebra in $\mathcal{V}[G_\sigma]$ for which the canonical homomorphism $\kappa\colon G_\sigma \to Aut(\mathbf{F})_\sigma$ is an isomorphism. Let $\mathcal{V}$ be the variety generated by $\mathbf{F}$, and let $\mathcal{M}$ be a minimal, locally finite, congruence permutable subvariety of $\mathcal{V}$ with strictly simple generator $\mathbf{A}$. Assume that $\dim(\mathbf{A}) = 2$. If P is the set of covers of 0 in the linear quasiorder of Theorem 4.10 (2), then every permutation of $X := (\{0\} \times P) \cup (P \times \{0\})$ is induced by an element of G. Hence, for $n = |X|$, there is a surjective homomorphism of pointed groups,*

$$\mu\colon G_\sigma \twoheadrightarrow (S_n)_\sigma$$

where S_n is the symmetric group on n letters and $\sigma \in S_n$ is a fixed point free involution.

Proof. According to Theorem 4.10 (2), X is a union of G-orbits under the action defined in line (4.1). This yields a representation $G \to S_X$. Fix an enumeration $\{(a_i, b_i) \mid 1 \leq i \leq n\}$ of the pairs in X, and use this to re-express this representation as $\mu\colon G \to S_n$. Our tasks are to show that μ is surjective and that $\mu(\sigma)$ is fixed point free (in which case we will take $\mu(\sigma) \in S_n$ to be the element of S_n denoted σ, so that μ is a pointed group homomorphism).

We must show that for any $\lambda \in S_n$ there exists $\alpha \in G$ such that $\alpha(a_i, b_i) = (a_{\lambda(i)}, b_{\lambda(i)})$ for all i. If $1 \in P$, where 1 is the largest element in the linear quasiorder of Theorem 4.10, then $P = \{1\}$, $X = \{(0, 1), (1, 0)\}$, $n = 2$, and $id, \sigma \in G$ induce the two distinct permutations of X. This proves the theorem in the case where $1 \in P$, so henceforth we assume that $1 \notin P$.

Claim 4.14. *If* $\mathbf{a} = (a_1, \ldots, a_n)$ *and* $\mathbf{b} = (b_1, \ldots, b_n)$*, then* $\{\mathbf{a}, \mathbf{b}\}$ *generates* $\mathbf{A}^n$*.*

Proof. Let S be the subuniverse of $\mathbf{A}^n$ generated by $\{\mathbf{a}, \mathbf{b}\}$. By Lemma 4.8, the fact that $\mathbf{A}$ is quasiprimal (Theorem 4.12) implies that if S is a proper subuniverse of A^n then there are coordinates $1 \leq i < j \leq n$ such that the projection $S_{ij} := \pi_{ij}(S)$ is a proper subuniverse of A^2. The subuniverse S_{ij} contains $\{(a_i, a_j), (b_i, b_j)\} = \{(0,0), (p,q)\}$ or $\{(0,p), (p',0)\}$ where $p, p', q \in P$ and $p \neq q$. Since $0 \notin P$, S_{ij} is a subdirect subuniverse of $\mathbf{A}^2$, hence is the graph of an automorphism of $\mathbf{A}$. If $\{(a_i, a_j), (b_i, b_j)\} = \{(0,0), (p,q)\}$, then this automorphism fixes 0 and moves p to q. But no nonidentity automorphism of $\mathbf{A}$ fixes 0, according to Theorem 4.10 (3). If $\{(a_i, a_j), (b_i, b_j)\} = \{(0,p), (p',0)\}$, then this automorphism moves 0 to p and p' to 0. But a nonidentity automorphism of $\mathbf{A}$ can only move 0 to 1, according to Theorem 4.10 (3), and $1 \notin P$. Thus $S = A^n$, as claimed. □

Choose and fix $\lambda \in S_n$. Let $\mathbf{c} = (a_{\lambda(1)}, \ldots, a_{\lambda(n)})$ and $\mathbf{d} = (b_{\lambda(1)}, \ldots, b_{\lambda(n)})$. Using the same argument as that for Claim 4.14, $\{\mathbf{c}, \mathbf{d}\}$ generates $\mathbf{A}^n$. This shows that item (1) of Theorem 4.2 holds, when $\mathbf{C} = \mathbf{D} = \mathbf{A}^n$.

Let $\phi \in \mathrm{Vect}(\mathbf{A})$ be a nonconstant vector inducing the linear quasiorder on A that is described in Theorem 4.10. Suppose that $\phi(0) = r$ and $\phi(P) = \{t\}$. Let $\mathcal{B} = \{\phi_0, \ldots, \phi_n\}$ be the explicit basis of $\mathrm{Vect}(\mathbf{A}^n)$ described in Corollary 4.11, namely ϕ_0 is constant and $\phi_i(\mathbf{x}) = \phi(x_i)$. Since $\phi_i(\mathbf{a}) + \phi_i(\mathbf{b}) = \phi(a_i) + \phi(b_i)$ and $(a_i, b_i) = (0,p)$ or $(p,0)$, we get that $\phi_i(\mathbf{a}) + \phi_i(\mathbf{b}) = \phi(0) + \phi(p) = r + t$ for all i. Similarly, $\phi_i(\mathbf{c}) + \phi_i(\mathbf{d}) = r + t$ for all i. This is enough to imply that

$$\phi(\mathbf{a}) + \phi(\mathbf{b}) = \phi(\mathbf{c}) + \phi(\mathbf{d})$$

holds for all $\phi \in \mathcal{B}$, hence for all $\phi \in \mathrm{Vect}(\mathbf{A}^n)$. This shows that item (2) of Theorem 4.2 holds. The conclusion guaranteed by that theorem is that there is an $\alpha \in G$ such that $(\mathbf{c}, \mathbf{d}) = \alpha(\mathbf{a}, \mathbf{b}) = (t_\alpha(\mathbf{a}, \mathbf{b}), t_{\sigma\alpha}(\mathbf{a}, \mathbf{b}))$. By comparing coordinates we get the middle equality in

$$\alpha(a_i, b_i) = (t_\alpha(a_i, b_i), t_{\sigma\alpha}(a_i, b_i)) = (c_i, d_i) = (a_{\lambda(i)}, b_{\lambda(i)}).$$

This proves that $\alpha(a_i, b_i) = (a_{\lambda(i)}, b_{\lambda(i)})$ holds for all i, as desired.

The fact that $\mu(\sigma)$ is fixed point free, equivalently that σ acts without fixed points on $X = (\{0\} \times P) \cup (P \times \{0\}) \subseteq A \times A$, is a consequence of the facts that $0 \notin P$ and σ interchanges coordinates of $A \times A$. □

5. Free Algebras and Their Automorphism Groups

At the end of the introduction we gave a brief sketch of our approach to the representability problem for G_σ. Recall that we must determine $\mathrm{Aut}(\mathbf{E})$ for $\mathbf{E} = \mathbf{F}_{\mathcal{M}}(x, y)$, where $\mathcal{M}$ is a minimal locally finite variety. We do not know how to do this for arbitrary minimal locally finite $\mathcal{M}$, but in this section we will determine $\mathrm{Aut}(\mathbf{E})$ for $\mathbf{E} = \mathbf{F}_{\mathcal{M}}(x, y)$ where $\mathcal{M}$ is a congruence permutable, minimal, locally

finite variety. The assumption of congruence permutability guarantees that $\mathcal{M}$ is generated by a strictly simple algebra $\mathbf{A}$ that is quasiprimal or affine with a 1-element subuniverse, and that $\mathbf{E}$ is a power of $\mathbf{A}$. We consider the case $\mathbf{E} \cong \mathbf{A}^1$ first.

Theorem 5.1. *Let $\mathbf{A}$ be a strictly simple algebra. $\mathbf{A}$ is a 2-generated free algebra if and only if $\mathbf{A}$ is idempotent and $Aut(\mathbf{A})$ acts sharply 2-transitively on A.*

Proof. Since $\mathbf{A}$ has no nontrivial proper subuniverses, it is generated by any two distinct elements. $\mathbf{A}$ is freely generated by $\{a, b\} \subseteq A$ if and only if every function $f\colon \{a, b\} \to A$ extends to an endomorphism of $\mathbf{A}$. Since $\mathbf{A}$ is simple, the extension is either constant or an automorphism, with the two cases distinguished according to whether $f(a) = f(b)$. All constant functions $f\colon \{a, b\} \to A$ have extensions to endomorphisms if and only if $\mathbf{A}$ is idempotent, and all nonconstant functions have extensions to automorphisms if and only if $\mathrm{Aut}(\mathbf{A})$ acts 2-transitively on $\mathbf{A}$. The action must be sharp, since otherwise some nonidentity automorphism of $\mathrm{Aut}(\mathbf{A})$ fixes two elements, and the set of all fixed points would be a nontrivial proper subuniverse of $\mathbf{A}$. □

Theorem 5.2. *Suppose that H acts sharply 2-transitively on a finite set A. Then H is a Frobenius group whose Frobenius kernel K is elementary abelian of order $|K| = |A| = p^t$ and of index $[H : K] = p^t - 1$ for some prime p and some positive integer t. K is the intersection of all nontrivial normal subgroups of H, and $C_H(K) = K$.*

In fact, either there is a finite field $\mathbb{F}_q$, $q = p^t$, such that H is a subgroup of the group of semilinear transformations

$$x \mapsto ax^\alpha + b, \quad a \in \mathbb{F}_q^\times,\ b \in \mathbb{F}_q,\ \alpha \in Aut(\mathbb{F}_q)$$

containing the subgroup K of all translations ($x \mapsto x + b$), or else H is one of seven exceptional groups for which $p^t = 5^2, 7^2, 11^2, 11^2, 23^2, 29^2$, or 59^2 (there are two exceptional groups for which $p^t = 11^2$).

If a 2-Sylow subgroup of H is generalized quaternion, then it is an 8-element quaternion group and H is one of the exceptional groups for which $p^t = 5^2, 11^2, 11^2, 29^2$, or 59^2.

Proof. The information of the first two paragraphs can be derived from Section XII.9 of [7], particularly Theorems 9.1 and 9.8. If H is not exceptional, then a 2-Sylow subgroup is a semidirect product of two cyclic groups, so it is not generalized quaternion. If H is exceptional, then it follows from Remark 9.5 of [7] that the 2-Sylow subgroup of H is the 8-element quaternion group Q_8 in the cases where $p^t = 5^2, 11^2, 11^2, 29^2$, or 59^2, and has the form $Q_8 \rtimes \mathbb{Z}_2$ in the cases where $p^t = 7^2$ or 23^2. □

Next we discuss the situation where $\mathbf{E} \cong \mathbf{A}^n$ and $n > 1$, considering the quasiprimal case first. Let $\mathbf{A}$ be a strictly simple quasiprimal algebra. Let $\mathrm{Aut}(\mathbf{A})$ act diagonally on $A \times A$ on the right, i.e. $(a, b)\alpha := (a\alpha, b\alpha)$. A *bad orbit* under this action is an orbit containing a pair (a, a) where $\{a\}$ is a 1-element subuniverse of $\mathbf{A}$. (If an orbit

has one pair of this type, then every pair in the orbit is of this type.) An orbit that is not bad is *good*. A subset of $A \times A$ will be called a *good set of pairs* if it contains exactly one pair from each good orbit.

Lemma 5.3. *Let* $\mathbf{A}$ *be a strictly simple quasiprimal algebra. If* $\{(a_i, b_i) \mid 1 \leq i \leq n\}$ *is a good set of pairs, then* $\mathbf{A}^n$ *is free on two generators in the variety generated by* $\mathbf{A}$ *and* $\mathbf{a} = (a_1, \ldots, a_n)$*, and* $\mathbf{b} = (b_1, \ldots, b_n)$ *are free generators of* $\mathbf{A}^n$*.*

Proof. Let S be the subuniverse of $\mathbf{A}^n$ generated by $\{\mathbf{a}, \mathbf{b}\}$. We use Lemma 4.8 to show that $S = A^n$. If this is not the case, then S is contained in a maximal proper subuniverse. If S is contained in a maximal subuniverse of the form $A \times \cdots \times \underbrace{\{0\}}_{i} \times \cdots \times A$ for some i and some 1-element subuniverse $\{0\}$, then $(a_i, b_i) = (0, 0)$. But this is not the case, since $(0, 0)$ represents a bad orbit and no pair (a_i, b_i) represents such an orbit. If S is contained in a maximal subuniverse of the form $\{(a_1, \ldots, a_n) \mid a_i = a_j\alpha \text{ for some } i \neq j \text{ and some } \alpha \in \mathrm{Aut}(\mathbf{A})\}$, then $(a_i, b_i) = (a_j, b_j)\alpha$. But this is not the case, since our set of good pairs contains only one pair from each $\mathrm{Aut}(\mathbf{A})$-orbit. Thus, $S = A^n$, indeed.

To prove that $\mathbf{A}^n$ is free over $\{\mathbf{a}, \mathbf{b}\}$, it will suffice to prove that $\mathbf{A}^n$ is the largest 2-generated algebra in the variety generated by $\mathbf{A}$. Suppose that $\mathbf{A}^m \in \mathsf{HSP}(\mathbf{A})$ is generated by $\mathbf{c} = (c_1, \ldots, c_m)$ and $(d_1, \ldots, d_m)$. We cannot have $(c_i, d_i) = (0, 0)$ for some 1-element subuniverse $\{0\}$ of $\mathbf{A}$, for then $\mathbf{c}$ and $\mathbf{d}$ both lie in the proper subuniverse of $\mathbf{A}^n$ consisting of tuples with 0 in the i-th coordinate. If $(c_i, d_i) = (c_j, d_j)\alpha$ for some $i \neq j$ and some $\alpha \in \mathrm{Aut}(\mathbf{A})$, then $\mathbf{c}$ and $\mathbf{d}$ both lie in the proper subuniverse consisting of all tuples $\mathbf{x}$ with $x_i = x_j\alpha$. Thus, if $\mathbf{A}^m$ is generated by $\{\mathbf{c}, \mathbf{d}\}$, then $\{(c_i, d_i) \mid 1 \leq i \leq m\}$ is a set of pairs from distinct good orbits. This shows that $\leq n$, since n is the number of good orbits. □

A semidirect product written $M \rtimes Q$ will consist of pairs $(m, q) \in M \times Q$ with multiplication given by $(m_1, q_1)(m_2, q_2) = (m_1\gamma(q_1)(m_2), q_1q_2)$ where $\gamma\colon Q \to \mathrm{Aut}(M)$ is a homomorphism. If written $Q \ltimes M$, then it consists of pairs $(q, m) \in Q \times M$ with multiplication given by $(q_1, m_1)(q_2, m_2) = (q_1q_2, \gamma(q_2)(m_1)m_2)$ where $\gamma\colon Q \to \mathrm{Aut}(M)^{op}$ is a homomorphism. A wreath product $M \wr S_n$ is a semidirect product $M^n \rtimes S_n$ where $\gamma\colon S_n \to \mathrm{Aut}(M^n)$ is defined by $\gamma(\alpha)(m_1, \ldots, m_n) = (m_{\alpha^{-1}(1)}, \ldots, m_{\alpha^{-1}(n)})$. We will typically write this as a right action and drop the reference to γ, i.e., $(m_1, \ldots, m_n)\alpha = (m_{\alpha(1)}, \ldots, m_{\alpha(n)})$,

Lemma 5.4. *If* $\mathbf{A}$ *is quasiprimal, then* $Aut(\mathbf{A}^n) = Aut(\mathbf{A}) \wr S_n$*. If* $\mathbf{A}$ *is strictly simple, then* $Aut(\mathbf{A})$ *acts on* A *so that no nonidentity element has more than one fixed point.*

Proof. The automorphisms of $\mathbf{A}^n$ include those in $\mathrm{Aut}(\mathbf{A}) \wr S_n$ for any algebra $\mathbf{A}$. Here $\alpha \in \mathrm{Aut}(\mathbf{A})$ acts on the right in the i-th coordinate by $(a_1, \ldots, a_i, \ldots, a_n) \mapsto (a_1, \ldots, a_i\alpha, \ldots, a_n)$ and $\lambda \in S_n$ acts on the right on the set of coordinates by

$(a_1, \dots, a_n) \mapsto (a_{\lambda(1)}, \dots, a_{\lambda(n)})$. These generate all automorphisms of the form $(a_1, \dots, a_n) \mapsto (a_{\lambda(1)}\alpha_1, \dots, a_{\lambda(n)}\alpha_n)$. What we must show is that there are no other automorphisms when $\mathbf{A}$ is quasiprimal. The property of quasiprimal algebras that we will use is that $\mathbf{A}^n$ has a unique direct factorization as an n-th power, which follows from the distributivity of $\mathbf{Con}(\mathbf{A}^n)$.

Let θ_i be the kernel of the i-th coordinate projection $\pi_i \colon \mathbf{A}^n \to \mathbf{A}$. If $\Gamma \colon \mathbf{A}^n \to \mathbf{A}^n$ is an automorphism, then the uniqueness of direct decompositions forces $\{(\theta_i)\Gamma^{-1}\} = \{\theta_i\}$, so there is a permutation $\lambda \in S_n$ such that $\theta_{\lambda(i)} = (\theta_i)\Gamma^{-1}$. By the first isomorphism theorem, the surjective homomorphism $\mathbf{A}^n \xrightarrow{\Gamma} \mathbf{A}^n \xrightarrow{\pi_i} \mathbf{A}$ can be factored as $\mathbf{A}^n \xrightarrow{\pi_j} \mathbf{A} \xrightarrow{\alpha_j} \mathbf{A}$ where $j = \lambda(i)$ and $\alpha_j \in \mathrm{Aut}(\mathbf{A})$, We have $(a_1, \dots, a_n)\Gamma \mapsto (a_{\lambda(1)}\alpha_1, \dots, a_{\lambda(n)}\alpha_n)$, so $\Gamma \in \mathrm{Aut}(\mathbf{A}) \wr S_n$, as desired.

That each $\alpha \in \mathrm{Aut}(\mathbf{A}) - \{1\}$ acts with at most one fixed point when $\mathbf{A}$ is strictly simple follows from the fact that the set of fixed points of α is a subuniverse. □

The diagonal action of $\mathrm{Aut}(\mathbf{A})$ on $A \times A$ commutes with the interchanging of coordinates. Therefore, the reflection of the orbit of a pair (u, v) across the diagonal of $A \times A$ is the orbit of the pair (v, u). Either both or neither of the orbits are good. We call a pair (u, v) from a good orbit *isolated* or *nonisolated* according to whether the orbits of (u, v) and (v, u) are disjoint or equal.

We will distinguish between *diagonal* nonisolated pairs (those pairs from a good orbit that have the form (u, u)) and *off-diagonal* nonisolated pairs (the rest of the nonisolated pairs). A *good sequence of pairs with parameters k, ℓ and m* is a sequence $\langle (a_i, b_i) \mid 1 \le i \le n \rangle$ where $\{(a_i, b_i) \mid 1 \le i \le n\}$ is a good set of pairs, and

(1) $(a_{2i-1}, b_{2i-1}) = (u, v)$ and $(a_{2i}, b_{2i}) = (v, u)$ are isolated pairs, for $1 \le i \le k$,
(2) $(a_{2k+i}, b_{2k+i}) = (u, u)$ is a diagonal nonisolated pair for $1 \le i \le \ell$,
(3) $(a_{2k+\ell+i}, b_{2k+\ell+i}) = (u, v)$ is an off-diagonal nonisolated pair for $1 \le i \le m$, and
(4) $2k + \ell + m = n$.

In other words, a good sequence is obtained from a good set by ordering the pairs so that the isolated pairs come first, followed by the diagonal nonisolated pairs, and then the off-diagonal nonisolated pairs. Moreover, we assume that isolated pairs are chosen *in pairs*, (u, v) followed by (v, u).

Theorem 5.5. *Let $\mathbf{A}$ be a strictly simple quasiprimal algebra. If $\langle (a_i, b_i) \mid 1 \le i \le n \rangle$ is a good sequence of pairs with parameters k, ℓ and m, then $\mathbf{A}^n$ is free on two generators in the variety generated by $\mathbf{A}$ and $\mathbf{a} = (a_1, \dots, a_n)$, and $\mathbf{b} = (b_1, \dots, b_n)$ are free generators of $\mathbf{A}^n$. The element of $Aut(\mathbf{A}^n)$ that interchanges $\mathbf{a}$ and $\mathbf{b}$ is the element $\sigma := (\alpha, \beta) \in Aut(\mathbf{A}) \wr S_n$ where*

(1) $\beta = (1\ 2)(3\ 4) \cdots (2k-1\ 2k) \in S_n$.

(2) $\alpha = (\alpha_1, \dots, \alpha_n) \in Aut(\mathbf{A})^n$ *is the automorphism where* $\alpha_i = id$ *for* $1 \leq i \leq k + \ell$ *and* α_j *is the unique involution in* $Aut(\mathbf{A})$ *satisfying* $(a_j, b_j)\alpha_j = (b_j, a_j)$ *for* $k + \ell < j \leq n$.

Proof. We first clarify the use of the word "unique" in the last sentence of the theorem statement. If $a \neq b$ and $\alpha, \beta \in \mathrm{Aut}(\mathbf{A})$ satisfy $(a, b)\alpha = (a, b)\beta = (b, a)$, then $\alpha\beta^{-1}$ fixes both a and b. But no nonidentity element of $\mathrm{Aut}(\mathbf{A})$ has more then one fixed point, so $\alpha = \beta$. Thus, there is at most one $\alpha \in \mathrm{Aut}(\mathbf{A})$ such that $(a, b)\alpha = (b, a)$, while if (a, b) is a nonisolated pair then there is at least one.

Since a good sequence is a special ordering of the pairs in a good set, the fact that $\mathbf{A}^n$ is freely generated by $\{\mathbf{a}, \mathbf{b}\}$ follows from Lemma 5.3, while the fact that $\mathrm{Aut}(\mathbf{A}^n) \cong \mathrm{Aut}(\mathbf{A}) \wr S_n$ follows from Lemma 5.4. Since $(a_{2i}, b_{2i}) = (b_{2i-1}, a_{2i-1})$ for $1 \leq i \leq k$ and $(a_i, b_i) = (a_i, a_i)$ for $2k < i \leq 2k + \ell$, the tuples $\mathbf{a}$ and $\mathbf{b}$ are

$$\begin{aligned}\mathbf{a} &= (a_1, b_1, a_3, b_3, \dots, a_{2k-1}, b_{2k-1}, a_{2k+1}, \dots, a_{2k+\ell}, a_{2k+\ell+1}, \dots, a_{2k+\ell+m})\\ \mathbf{b} &= (b_1, a_1, b_3, a_3, \dots, b_{2k-1}, a_{2k-1}, a_{2k+1}, \dots, a_{2k+\ell}, b_{2k+\ell+1}, \dots, b_{2k+\ell+m}).\end{aligned}$$

If we apply σ to $\mathbf{a}$ we get

$$\begin{aligned}\mathbf{a}\sigma &= (a_{\beta(1)}, b_{\beta(1)}, \dots, a_{2k+1}, \dots, a_{2k+\ell}, a_{2k+\ell+1}\alpha_{2k+\ell+1}, \dots, a_{2k+\ell+m}\alpha_{2k+\ell+m})\\ &= (a_2, b_2, \dots, a_{2k+1}, \dots, a_{2k+\ell}, b_{2k+\ell+1}, \dots, b_{2k+\ell+m})\\ &= (b_1, a_1, \dots, a_{2k+1}, \dots, a_{2k+\ell}, b_{2k+\ell+1}, \dots, b_{2k+\ell+m}) = \mathbf{b},\end{aligned}$$

and similarly $\mathbf{b}\sigma = \mathbf{a}$. □

Now we consider the case where $\mathbf{A}$ is affine.

Lemma 5.6. *A strictly simple affine algebra* $\mathbf{A}$ *with a* 1*-element subuniverse* $\{0\}$ *is term equivalent to an algebra of the following type. There is a finite field* $\mathbb{F}_q$ *and parameters* $0 \leq d \leq m$ *such that the universe is* $A = \mathbb{F}_q^m$ *and the term operations are precisely the linear operations of the form* $M_1x_1 + \cdots + M_kx_k$ *where* $M_i \in M_m(\mathbb{F}_q)$ *and* $M := \sum M_i$ *is a matrix for which* $Me_i = e_i$ *for* $1 \leq i \leq d$. *Here* e_i *is the* i*-th standard basis vector.*

The automorphisms of such an algebra are the functions of the form $\lambda x + c$ *where* $\lambda \in \mathbb{F}_q^\times$ *and* $c \in A$ *belongs to the subspace* V *that is generated by* $\{e_i \mid 1 \leq i \leq d\}$.

Proof. The claims of the first paragraph follow from Proposition 2.6 of [9], which fully describes the clones of all affine algebras. (The claims can be derived even more easily from Proposition 2.10 of [9], which specializes the result to strictly simple affine algebras). Here V is the set of elements of A that are 1-element subuniverses.

If $\alpha \in \mathrm{Aut}(\mathbf{A})$, then $0\alpha = c$ for some $c \in V$. It follows from the description of the clone of $\mathbf{A}$ that the translation $T_{-c}\colon x \mapsto x - c$ is an automorphism of $\mathbf{A}$, so $T_{-c} \circ \alpha$ is an automorphism of $\mathbf{A}$ that fixes 0. Hence $T_{-c} \circ \alpha$ is an automorphism of A considered as a module over $M_m(\mathbb{F}_q)$, so $T_{-c} \circ \alpha = \lambda x$ for some $\lambda \in \mathbb{F}_q^\times$. Therefore, $\alpha = T_c(\lambda x) = \lambda x + c$. □

Henceforth we will assume that any strictly simple affine algebra with a 1-element subuniverse has the form described in the theorem, and that the meaning of the notation $\mathbb{F}_q, m, d, V$ etc. is understood.

Lemma 5.7. *Let* $\mathbf{A}$ *be a strictly simple affine algebra with a* 1*-element subuniverse. If* S *is a maximal proper subuniverse of* $\mathbf{A}^n$ *for some* $n \geq 1$*, then there exist* $\lambda_i \in \mathbb{F}_q$*, not all zero, and* $c \in V$ *such that*

$$S = \left\{ (a_1, \ldots, a_n) \;\middle|\; \sum_{i=1}^{n} \lambda_i a_i = c \right\}. \tag{5.1}$$

Proof. Every subuniverse of $\mathbf{A}^n$ is a congruence class, and every congruence class containing a subuniverse is itself a subuniverse. Since the congruences of an affine algebra are uniform, a maximal subuniverse is a class of a unique maximal congruence. Let θ be the congruence that has S as a class. Then $\mathbf{A}^n/\theta \cong \mathbf{A}$, so there is a surjective homomorphism $\varphi\colon \mathbf{A}^n \to \mathbf{A}$ whose kernel is θ. If $\mathbf{0} = (0, \ldots, 0)$, then by composing φ with the translation $T_{-\varphi(\mathbf{0})}$ if necessary, we may assume that $\varphi(\mathbf{0}) = 0$, i.e., φ is a surjective module homomorphism. Moreover, $\varphi(S) = \{c\}$ is a 1-element subuniverse of $\mathbf{A}$, so $S = \varphi^{-1}(c)$ for some $c \in V$.

Any module homomorphism $\varphi \in \mathrm{Hom}(\mathbf{A}^n, \mathbf{A})$ is an n-ary operation of the centralizer of the clone of $\mathbf{A}$, which is the clone of a module over the ring $\mathrm{End}(\mathbf{A})$. Hence φ has the form $\sum_{i=1}^{n} \varepsilon_i(x_i)$ where $\varepsilon_i \in \mathrm{End}(\mathbf{A})$. Since the module endomorphisms of $\mathbf{A}$ are the functions $x \mapsto \lambda x$, $\lambda \in \mathbb{F}_q^\times$, and φ is surjective, $\varphi(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i x_i$ where $\lambda_i \in \mathbb{F}_q^\times$ and not all λ_i are zero. Thus, if S is a maximal subuniverse of $\mathbf{A}^n$, then $S = \varphi^{-1}(c) = \{(a_1, \ldots, a_n) \mid \sum_{i=1}^{n} \lambda_i a_i = c\}$ where $c \in V$ and not all λ_i are zero. □

If $\mathbf{A}$ is a strictly simple affine algebra with a 1-element subuniverse, then $\mathbb{F}_q$ acts on $A \times A$ diagonally making this set a $2m$-dimensional $\mathbb{F}_q$-vector space. Write Δ_V for the subspace consisting of all pairs $(c, c), c \in V$. A *good set of pairs* for $\mathbf{A}$ is a subset $\{(a_1, b_i) \mid 1 \leq i \leq n\} \subseteq A \times A$ which is a basis modulo Δ_V. That is, it is a maximal subset of $A \times A$ such that $\sum_{i=1}^{n} \lambda_i(a_i, b_i) = (c, c) \in \Delta_V$ implies $\lambda_i = 0$ for all i.

Lemma 5.8. *Let* $\mathbf{A}$ *be a strictly simple affine algebra with a* 1*-element subuniverse and parameters* $d \leq m$*. The size of any good set of pairs is* $n = 2m - d$*. If* $\{(a_i, b_i) \mid 1 \leq i \leq n\}$ *is a good set of pairs, then* $\mathbf{A}^n$ *is free on two generators in the variety generated by* $\mathbf{A}$ *and* $\mathbf{a} = (a_1, \ldots, a_n)$*, and* $\mathbf{b} = (b_1, \ldots, b_n)$ *are free generators of* $\mathbf{A}^n$*.*

Proof. The size of a good set of pairs is the $\mathbb{F}_q$-dimension of $(A \times A)/\Delta_V$, which is $2m - d$ since the dimension of $\mathbf{A}$ is m and the dimension of Δ_V equals $\dim(V) = d$.

The set $\{(a_i, b_i) \mid 1 \leq i \leq n\}$ fails to be independent modulo Δ_V if and only if there is a dependence relation $\sum_{i=1}^{n} \lambda_i(a_i, b_i) = (c, c) \in \Delta_V$ where not all λ_i are zero. This holds if and only if the tuples $\mathbf{a}$ and $\mathbf{b}$ belong to the maximal proper subuniverse

defined by

$$\left\{(x_1,\dots,x_n) \;\middle|\; \sum_{i=1}^{n}\lambda_i x_i = c\right\}.$$

Thus, $\{(a_i,b_i) \mid 1\le i\le n\}$ fails to be independent modulo Δ_V if and only if $\{\mathbf{a},\mathbf{b}\}$ fails to generate $\mathbf{A}^n$. It follows that $\mathbf{A}^n$ is freely generated by $\{\mathbf{a},\mathbf{b}\}$ if and only if $\{(a_i,b_i) \mid 1\le i\le n\}$ is a good set of pairs. □

Theorem 5.9. *Let* $\mathbf{A}$ *be a strictly simple affine algebra with a* 1*-element subuniverse and parameters* $d\le m$*. The set*

$$Z := \{(e_1,0),(e_2,0),\dots,(e_d,0),(e_{d+1},0),(0,e_{d+1}),\dots,(e_m,0),(0,e_m)\}$$

is a good set of pairs. The algebra $\mathbf{A}^n = \mathbf{A}^{2m-d}$ *is free on two generators in the variety generated by* $\mathbf{A}$*, and*

$$\mathbf{a} := (e_1,e_2,\dots,e_d,e_{d+1},0,e_{d+2},0,\dots,e_m,0)$$

and

$$\mathbf{b} := (0,0,\dots,0,0,e_{d+1},0,e_{d+2},\dots,0,e_m)$$

are free generators. Aut$(\mathbf{A}^n)$ *consists of all linear functions of the form* $L\mathbf{x}+K$ *where* $L=[\lambda_{ij}]\in \mathrm{GL}(n,q)$ *and* $K=(c_1,\dots,c_n)^t\in V^n$*, and is therefore isomorphic to* $\mathrm{GL}(n,q)\ltimes T$ *where* T *is an elementary abelian group of order* q^{dn}*. The automorphism* $\sigma\in$ *Aut*$(\mathbf{A}^n)$ *that interchanges* $\mathbf{a}$ *and* $\mathbf{b}$ *is*

$$L\mathbf{x}+K = \begin{bmatrix} -I & 0 & 0 & 0 & 0\\ 0 & S & 0 & 0 & 0\\ 0 & 0 & S & 0 & 0\\ \vdots & & & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & S \end{bmatrix}\begin{bmatrix} X_0\\ X_1\\ X_2\\ \vdots\\ X_{m-d}\end{bmatrix} + \begin{bmatrix}\mathbf{c}\\0\\0\\\vdots\\0\end{bmatrix}$$

where $-I$ *is the negative of the* $d\times d$ *identity matrix,* $S=\begin{bmatrix}0&1\\1&0\end{bmatrix}$, $X_0=(x_1,\dots,x_d)^t$, $X_i=(x_{d+2i-1},x_{d+2i})^t$ *if* $i>1$*, and* $\mathbf{c}=(e_1,\dots,e_d)^t$.

Proof. The set Z has $2m-d$ elements, so it is the right size to be a good set of pairs. It suffices to show that it is independent modulo Δ_V. Choose any dependence relation $\sum_{i=1}^n \lambda_i(a_i,b_i)=(c,c)\in\Delta_V$ where $\mathbf{a}=(a_1,\dots,a_n)$ and $\mathbf{b}=(b_1,\dots,b_n)$ are defined as in the theorem statement. Considering this dependence relation in the second coordinate only yields $\sum_{i=1}^n\lambda_i b_i=\sum_{j=1}^{m-d}\lambda_{d+2j}e_{d+j}=c$. But $c\in V$ and $\{e_{d+1},\dots,e_m\}$ spans a space complementary to V. Thus $c=0$ and $\lambda_{d+2j}=0$ for $1\le j\le m-d$. Now considering the dependence relation in the first coordinate yields $\sum_{i=1}^n\lambda_i a_i=\sum_{j=1}^d\lambda_j a_j+\sum_{k=1}^{m-d}\lambda_{d+2k-1}e_{d+k}=c=0$. This is a $\mathbb{F}_q$-dependence relation among the basis vectors $\{e_1,\dots,e_m\}$, hence $\lambda_j=0$ when $1\le j\le d$ and

$\lambda_{d+2k-1} = 0$ when $1 \leq k \leq m - d$. This accounts for all λ's, proving that the original dependence relation was trivial. Hence Z is independent modulo Δ_V.

The 1-element subuniverses of $\mathbf{A}^n$ are the elements $K \in V^n$, so the translations $T_K \colon \mathbf{x} \mapsto \mathbf{x}+K$ are among the automorphisms of $\mathbf{A}^n$. The group $T \cong \langle V; +, -, 0\rangle^n$ of all translations is an elementary abelian group of order $|V|^n = (q^d)^n$. If $\alpha \in \mathrm{Aut}(\mathbf{A}^n)$, then $K := \mathbf{0}\alpha$ is a 1-element subuniverse, and $T_{-K} \circ \alpha$ is an automorphism of $\mathbf{A}^n$ that fixes $\mathbf{0}$, i.e., it is a module automorphism. But the group of module automorphisms of a direct sum of n copies of a simple module $\mathbf{A}$ over $\mathbb{F}_q$ is just $\mathrm{GL}(n, q)$ acting on $\mathbf{A}^n$ (considered as a module of column vectors) by left multiplication. Thus, $\mathbf{x}(T_{-K} \circ \alpha) = L\mathbf{x}$ for some $L \in \mathrm{GL}(n, q)$, and therefore $\mathbf{x}\alpha = L\mathbf{x} + K \in \mathrm{GL}(n, q) \ltimes T$.

It remains to check that σ is represented by $L\mathbf{x} + K$ where L and K are as stated. For this, one must show that $\mathbf{b} = L\mathbf{a} + K$ and $\mathbf{a} = L\mathbf{b} + K$, which is too obvious for words. □

6. Main Results

In this section we draw upon our previous work to prove that certain pointed groups do not arise as the automorphism group of a finite, 2-generated free algebra.

We will derive results concerning a diverse collection of pointed groups, but the scheme of proof is the same in all instances. We will start with a finite pointed group G_σ that forces congruence permutability. If G_σ is representable as the automorphism group of a *finite*, 2-generated, free algebra $\mathbf{F}$, then according to Theorem 2.3 we may assume that $\mathbf{F}$ belongs to $\mathcal{V}[G_\sigma]$ and the canonical homomorphism $\kappa \colon G_\sigma \to \mathrm{Aut}(\mathbf{F})_\sigma$ is an isomorphism. Let $\mathcal{V}$ be the variety generated by $\mathbf{F}$, and let $\mathcal{M}$ be a minimal subvariety of $\mathcal{V}$. Since $\mathcal{V}$ is congruence permutable, so is $\mathcal{M}$, hence $\mathcal{M}$ is generated by a strictly simple algebra $\mathbf{A}$ that is either quasiprimal or affine with a 1-element subuniverse.

Let $\mathbf{E} = \mathbf{F}_{\mathcal{M}}(x_{\mathbf{E}}, y_{\mathbf{E}})$ be the 2-generated free algebra in $\mathcal{M}$. The natural surjective homomorphism $\nu \colon \mathbf{F} \to \mathbf{E} \colon x_{\mathbf{F}} \mapsto x_{\mathbf{E}}, y_{\mathbf{F}} \mapsto y_{\mathbf{E}}$ induces a homomorphism

$$\widehat{\nu} \colon \mathrm{Aut}(\mathbf{F}) \to \mathrm{Aut}(\mathbf{E}) \colon \alpha \mapsto \nu \circ \alpha \circ \nu^{-1}. \tag{6.1}$$

(That $\widehat{\nu}(\alpha)$ is a well-defined function follows from the fact that $\ker(\nu)$ is a characteristic congruence of $\mathbf{F}$.) Here ν acts on the left, but α acts on the right, so what is meant in (6.1) is that $(a)\widehat{\nu}(\alpha) = \nu((\nu^{-1}(a))\alpha)$.

We would like to understand when an automorphism $\gamma \in \mathrm{Aut}(\mathbf{E})$ is in the image of $\widehat{\nu}$. If $\gamma = \widehat{\nu}(\alpha)$, then

$$(x_{\mathbf{E}}, y_{\mathbf{E}})\gamma = (x_{\mathbf{E}}, y_{\mathbf{E}})\widehat{\nu}(\alpha) = (t_\alpha(x_{\mathbf{E}}, y_{\mathbf{E}}), t_{\sigma\alpha}(x_{\mathbf{E}}, y_{\mathbf{E}})).$$

Thus, $\gamma \in \mathrm{im}(\widehat{\nu})$ if and only if the pair $(x_{\mathbf{E}}\gamma, y_{\mathbf{E}}\gamma) \in E \times E$ lies in the same G-orbit as $(x_{\mathbf{E}}, y_{\mathbf{E}})$ under the action defined in line (4.1). By Theorem 4.2, this will be true if two conditions hold:

(1) $\{x_{\mathbf{E}}, y_{\mathbf{E}}\}$ and $\{x_{\mathbf{E}}\gamma, y_{\mathbf{E}}\gamma\}$ generate the same subalgebra of $\mathbf{E}$.

In fact, $\mathrm{Sg}^{\mathbf{E}}(\{x_{\mathbf{E}}, y_{\mathbf{E}}\}) = \mathrm{Sg}^{\mathbf{E}}(\{x_{\mathbf{E}}\gamma, y_{\mathbf{E}}\gamma\}) = \mathbf{E}$, since $\{x_{\mathbf{E}}, y_{\mathbf{E}}\}$ is a free generating set and γ is an automorphism. Thus, the second is the only interesting condition:

(2) $\phi(x_{\mathbf{E}}) + \phi(y_{\mathbf{E}}) = \phi(x_{\mathbf{E}}\gamma) + \phi(y_{\mathbf{E}}\gamma)$ for all $\phi \in \mathrm{Vect}(\mathbf{E})$.

Whether or not condition (2) holds can be determined with the aid of Theorem 4.9 and Corollary 4.11, since $\mathbf{E} \cong \mathbf{A}^n$. In particular, if $\dim(\mathbf{A}) = 1$, then $\dim(\mathbf{E}) = 1$ according to Theorem 4.9. In this case, $\phi(x_{\mathbf{E}}) + \phi(y_{\mathbf{E}}) = \phi(x_{\mathbf{E}}\gamma) + \phi(y_{\mathbf{E}}\gamma)$ holds trivially for all $\phi \in \mathrm{Vect}(\mathbf{E})$, since all ϕ are constant. Thus, when $\dim(\mathbf{A}) = 1$, the homomorphism $\widehat{\nu}$ is surjective. When $\dim(\mathbf{A}) = 2$, then $\dim(\mathbf{E}) \neq 1$ and $\widehat{\nu}$ is not surjective. Its image could be determined completely using Corollary 4.11 provided we had enough information about $\mathbf{A}$ (e.g., if we knew a nonconstant $\phi \in \mathrm{Vect}(\mathbf{A})$ and the exponent in the equation $\mathbf{E} \cong \mathbf{A}^n$). We avoid the problem of determining $\mathrm{im}(\widehat{\nu})$ by appealing to Theorem 4.13. If $\widehat{\nu}$ is not surjective, then this theorem produces a surjective homomorphism μ that can be used instead. Altogether, the conclusion is:

Theorem 6.1. *Let G be a finite group with designated involution σ. If G_σ forces congruence permutability and is representable as $Aut(\mathbf{F})_\sigma$ for some finite, 2-generated free algebra $\mathbf{F}$, then there is a surjective homomorphism $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ where H_σ is a pointed group of a type described in Theorems 4.13, 5.5 or 5.9. (That is, either $H = S_n$ and σ is a fixed point free permutation, $H = Aut(\mathbf{A}) \wr S_n$ and σ is as is described in Theorem 5.5, or $H = \mathrm{GL}(n, q) \ltimes T$ and σ is as is described in Theorem 5.9).*

This result is the main theorem of this paper. It is worth observing that, conversely, if there is an *isomorphism* $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ where H_σ is one of the three types of pointed groups, then G_σ is representable as $\mathrm{Aut}(\mathbf{F})_\sigma$ for some finite, 2-generated free algebra $\mathbf{F}$. That S_n with a fixed point free involution is representable was established in Theorem 2.16. The other two types are representable because they were originally identified as automorphism groups of finite, 2-generated free algebras in Theorems 5.5 and 5.9.

We proceed to make the general statement of Theorem 6.1 specific by showing that many of the groups described in the examples of Section 2 are not representable as $\mathrm{Aut}(\mathbf{F}_{\mathcal{V}}(x, y))_\sigma$ when $\mathbf{F}_{\mathcal{V}}(x, y)$ is finite.

Lemma 6.2. *Let G be a finite group with involution σ. If G has no retraction onto $\langle\sigma\rangle$, then every finite algebra $\mathbf{B} \in \mathcal{V}[G_\sigma]$ has cardinality $|B| \equiv 0$ or $1 \pmod 4$.*

Proof. The action of G on $B \times B$ defined in line (4.1) yields a homomorphism $\varphi\colon G \to S_{|B|^2}$ in which σ (which switches coordinates in $B \times B$) maps to a product of $|B|^2 - |B|$ disjoint transpositions. If $|B| \not\equiv 0, 1 \pmod 4$, then $|B|^2 - |B|$ is odd, and so σ maps to an odd permutation. Composing $\varphi\colon G \to S_{|B|^2}$ with the sign homomorphism $S_{|B|^2} \to \{\pm 1\}$ yields a homomorphism whose kernel is a normal complement N to $\langle\sigma\rangle$ in G. But then N is the kernel of a retraction of G onto $\langle\sigma\rangle$. □

Theorem 6.3. *Let G be a finite group with designated involution σ. If $C := C_G(\sigma)$ is a normal Hall subgroup of G and C has no retraction onto $\langle\sigma\rangle$, then G_σ is not representable as $Aut(\mathbf{F})_\sigma$ for any finite, 2-generated, free algebra $\mathbf{F}$.*

Proof. Assume otherwise that $G_\sigma \cong \mathrm{Aut}(\mathbf{F})_\sigma$ where $\mathbf{F}$ is a finite, 2-generated, free algebra. We may assume that $\mathbf{F}$ is free in a subvariety of $\mathcal{V}[G_\sigma]$. The fact that C has no retraction onto $\langle\sigma\rangle$ implies that G has no such retraction, so by Lemma 6.2 every finite algebra $\mathbf{B} \in \mathcal{V}[G_\sigma]$ has cardinality $|B| \equiv 0$ or $1 \pmod 4$.

According to Example 3.1, G_σ forces congruence permutability. Therefore there is a surjective homomorphism $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ where H_σ is a pointed group of a type described in Theorems 4.13, 5.5 or 5.9.

Let $D = C_G(C)$ be the double centralizer of σ. Then C and D are both normal in G, $\sigma \in D \leq C$, $[C, D] = \{id\}$, and C is Hall subgroup of G that has no retraction onto $\langle\sigma\rangle$. If $M = \varphi(C)$ and $N = \varphi(D)$, then

(1) M and N are both normal in H,
(2) $\sigma \in N \leq M$,
(3) $[M, N] = \{id\}$, and
(4) M is Hall subgroup of H,

since φ is surjective and preserves σ. It follows from these properties that

(5) M contains every element of H whose order is a power of 2,

since M is a normal Hall subgroup containing an involution. Finally,

(6) M has no retraction onto $\langle\sigma\rangle$,

since conjugating such a retraction by φ would lead to a retraction of C onto $\langle\sigma\rangle$. Fix this meaning for M and N for the remainder of the proof.

Case 1. $H = S_n$ and σ is a fixed point free involution.

Since M contains every element whose order is a power of 2, and $H = S_n$ is generated by involutions, $M = H = S_n$. Since $[M, N] = \{id\}$, N is a nontrivial central subgroup of $M = S_n$. Hence $n = 2$, in which case the identity function is a retraction of $H = M = N = \langle\sigma\rangle$ onto $\langle\sigma\rangle$, a contradiction. Thus, Case 1 cannot arise.

Case 2. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple quasiprimal algebra $\mathbf{A}$ (so $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$ and σ is as is described in Theorem 5.5).

Assume first that $n > 1$. The subgroup $S \leq H = \mathrm{Aut}(\mathbf{A}) \wr S_n$ consisting of pairs (α, β) with $\alpha = (id, \ldots, id) \in \mathrm{Aut}(\mathbf{A})^n$ is isomorphic to S_n, hence is generated by involutions. Therefore $S \leq M$, since M contains every element of H whose order is a power of 2, and consequently $N \leq C_H(M) \leq C_H(S)$. But when $n > 1$ the centralizer of S in $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$ is either

(i) the set of pairs (α, id) where $\alpha = (\tau, \tau, \ldots, \tau) \in \mathrm{Aut}(\mathbf{A})^n$ (when $n > 2$), or

(ii) the set of pairs (α, β) where $\alpha = (\tau, \tau) \in \mathrm{Aut}(\mathbf{A})^2$ and $\beta \in S_2$ (when $n = 2$).

For the normal subgroup N to consist of pairs of this type, τ must lie in the center of $\mathrm{Aut}(\mathbf{A})$. Otherwise, if $\tau^\delta := \delta^{-1}\tau\delta \neq \tau$, then the conjugate of $(\alpha, \beta) = ((\tau, \tau, \ldots, \tau, \tau), \beta) \in N$ by $((id, id, \ldots, id, \delta), id)$ is $((\tau, \tau, \ldots, \tau, \tau^\delta), \beta) \notin N$.

Now that we know the form of all elements of N, it is clear that form of the involution $\sigma \in N$ is $\sigma = ((\tau, \ldots, \tau), \beta)$ where either

(a) τ is a central involution in $\mathrm{Aut}(\mathbf{A})$, or
(b) $\tau = id$, $n = 2$ and $\beta = (1\ 2) \in S_2$.

If it is (a) that holds, then by Theorem 5.5 this form for σ implies that $A \times A$ has no isolated pairs nor any diagonal nonisolated pairs. Moreover, any off-diagonal $\mathrm{Aut}(\mathbf{A})$-orbit contains a pair (u, v) such that $(u, v)\tau = (u\tau, v\tau) = (v, u)$. But since $\tau \in Z(\mathrm{Aut}(\mathbf{A}))$, if this holds for one pair from every off-diagonal orbit, it holds for every pair from every off-diagonal orbit. Thus, $(u, v)\tau = (v, u)$ whenever $u \neq v$ in $\mathbf{A}$. This is impossible if $|A| > 2$. (If u, v and w are distinct, then $(u, v)\tau = (v, u)$ and $(u, w)\tau = (w, u)$, since the first equality implies $u\tau = v$ while the second implies that $u\tau = w$.) But since $|A| > 1$ and Lemma 6.2 guarantees that $|A| \equiv 0, 1 \pmod 4$, we do have $|A| > 2$. This is contradiction.

Now assume that it is (b) that holds, namely $n = 2$ and $\sigma = ((id, id), (1\ 2))$. For $\gamma \in \mathrm{Aut}(\mathbf{A})$, and $\Gamma := ((id, \gamma), id) \in H$, the element $\Lambda := \Gamma^{-1}\sigma\Gamma\sigma^{-1} = ((\gamma, \gamma^{-1}), id)$ belongs to the normal subgroup generated by σ. Since N is a normal abelian subgroup containing σ, N contains Λ, and therefore $\Lambda = ((\gamma, \gamma^{-1}), id)$ has the form $((\tau, \tau), \beta)$. Hence $\gamma = \gamma^{-1}$ for any $\gamma \in \mathrm{Aut}(\mathbf{A})$. This means that $\mathrm{Aut}(\mathbf{A})$ is an elementary abelian 2-group. The subgroup L of H consisting of all elements of the form $((\gamma, \gamma), \beta) \in \mathrm{Aut}(\mathbf{A}) \wr S_2$ is generated by σ and all elements of the form Λ above. Thus L is contained in the normal subgroup generated by σ, and therefore

$$L \leq N \leq M \leq C_H(\sigma) = L.$$

This proves that $M = L$ is the normal subgroup generated by σ, which is an elementary abelian 2-group. This contradicts the fact that M has no retraction onto $\langle\sigma\rangle$, thereby completing the proof that $n \not> 1$ in Case 2.

Now assume that $n = 1$. Then $\mathbf{A}$ is both strictly simple and free on two generators. By Theorem 5.1, $H = \mathrm{Aut}(\mathbf{A})$ acts sharply 2-transitively on A. Moreover, H has subgroups M and N satisfying the properties listed in the third paragraph of this proof. But no sharply 2-transitive group has subgroups satisfying all of these properties. The normal subgroup of H generated by σ must be an elementary abelian 2-group, since $\sigma \in N$ and N is normal and abelian. But since the Frobenius kernel K of H is the smallest nontrivial normal subgroup, it is contained in the normal subgroup generated by σ, so K must be an elementary abelian, normal, 2-Sylow subgroup of H. Thus $K \leq N \leq M$ and $M \leq C_H(N) \leq C_H(K) = K$. But then $M = K$ has

a retraction onto σ, because it is an elementary abelian 2-group containing σ. This concludes the proof that Case 2 cannot arise.

Case 3. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple affine algebra $\mathbf{A}$ with a 1-element subuniverse (so $H = \mathrm{GL}(n,q) \ltimes T$ and σ is as is described in Theorem 5.9).

Suppose that the parameters associated with $\mathbf{A}$ are $0 \leq d \leq m$, so that $n = 2m - d$. Consider first the situation where $n > 1$. M contains every element of $H = \mathrm{GL}(n,q) \ltimes T$ whose order is a power of 2. When $n > 1$ and q is even, this means that M contains $\mathrm{SL}(n,q) \ltimes T$. But then

$$\sigma \in N \leq C_H(M) \leq C_H(\mathrm{SL}(n,q) \ltimes T) = \{id\},$$

which is impossible. If $n > 1$ and q is odd, then the fact that M contains every element whose order is a power of 2 implies that M contains $\mathrm{SL}(n,q) \ltimes \{id\}$. In this situation

$$\sigma \in N \leq C_H(M) \leq C_H(\mathrm{SL}(n,q) \ltimes \{id\}) = Z(\mathrm{GL}(n,q)) \ltimes \{id\}.$$

Thus, $\mathbf{x}\sigma = -\mathbf{x}$, which can only happen when $d = m = n$, according to Theorem 5.9. Moreover, we have $M \leq C_H(N) \leq C_H(\sigma) = \mathrm{GL}(n,q) \ltimes \{id\}$. Thus q divides $|M|$, since M contains $\mathrm{SL}(n,q) \ltimes \{id\}$ and $n > 1$, and q also divides $|T| = q^{nd} = q^{n^2}$, which divides $[H : M]$. This contradicts the fact that M is a Hall subgroup. Thus, the assumption that $n > 1$ leads to a contradiction.

If $n = 1$, then $\mathbf{A}$ is a strictly simple algebra that is free on 2 generators. By Theorem 5.1, $\mathbf{A}$ is idempotent and $H = \mathrm{Aut}(\mathbf{A})$ acts sharply 2-transitively on A. Arguing as we did in the second paragraph of Case 2, we see that this cannot occur.

We have excluded all potential candidates for H_σ. The conclusion is G_σ is not representable. □

Corollary 6.4. *If $\sigma \in Z(G)$, then $G_\sigma \cong Aut(\mathbf{F})_\sigma$ for a finite, 2-generated, free algebra if and only if G has a retraction onto $\langle\sigma\rangle$.*

Proof. "If" is from Theorem 2.16; "only if" is from Theorem 6.3. □

Remark 6.5. The 8-element dihedral group is the underlying group for exactly two isomorphism types of pointed groups $(D_4)_\sigma$, σ an involution. The types can be distinguished according to whether $\sigma \in Z(D_4)$. The preceding theorem implies that $(D_4)_\sigma$ is not representable as $\mathrm{Aut}(\mathbf{F}_\mathcal{V}(x,y))_\sigma$ when $\sigma \in Z(D_4)$. But $(D_4)_\sigma$ is representable when $\sigma \notin Z(D_4)$, since this pointed group can be realized as the automorphism group of the free 2-generated rectangular band.

Theorem 6.6. *Let $\mathbf{F}$ be a finite, 2-generated, free algebra. If $G = Aut(\mathbf{F})$ has a 2-Sylow subgroup that is a generalized quaternion group or cyclic of order at least 4, then there is a surjective homomorphism $\varphi\colon G \twoheadrightarrow H$ where*

(1) *H is sharply 2-transitive.*
(2) *φ restricts to an isomorphism from any 2-Sylow subgroup of G onto a 2-Sylow subgroup of H.*

Proof. By Example 3.3, if the 2-Sylow subgroups of G are generalized quaternion or cyclic of order at least 4, then G_σ forces congruence permutability. Thus, there is a surjective homomorphism $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ where H_σ is of a type described in Theorems 4.13, 5.5 or 5.9. Since all involutions in G are conjugate, and $\varphi(\sigma) = \sigma \neq id$, it follows that the kernel of φ has odd order. Hence the restriction of φ to any 2-Sylow subgroup of H maps it isomorphically onto a 2-Sylow subgroup of H. This shows that the 2-Sylow subgroups of H are also generalized quaternion or cyclic of order at least 4. What remains to show is that this can only happen when H is sharply 2-transitive.

A criterion we will use to imply that a group has 2-Sylow subgroups that are neither generalized quaternion nor cyclic is that the group contains a pair of distinct commuting involutions. Then, each 2-Sylow subgroup will contain a pair of commuting involutions, but generalized quaternion groups and cyclic groups contain only one involution.

Case 1. $H = S_n$ and σ is a fixed point free involution.

If $n < 4$, then the 2-Sylow subgroup has order 2, and is therefore too small to be generalized quaternion or cyclic of order at least 4. If $n \geq 4$, then S_n contains a pair of commuting involutions, e.g., (1 2) and (3 4).

Case 2. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple quasiprimal algebra $\mathbf{A}$ (so $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$).

If $n = 1$, then H is sharply 2-transitive, and we are done. Thus we may assume that $n > 1$. If $\mathrm{Aut}(\mathbf{A})$ has even order, and $\tau \in \mathrm{Aut}(\mathbf{A})$ is an involution, then $((\tau, \tau, \ldots, \tau), id)$ and $((id, id, \ldots, id), (1\ 2))$ are distinct commuting involutions. In this case, the 2-Sylow subgroups of H cannot be generalized quaternion or cyclic of order at least 4. If $\mathrm{Aut}(\mathbf{A})$ has odd order, then the 2-Sylow subgroups of $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$ are isomorphic to those of S_n, and therefore cannot be generalized quaternion or cyclic of order at least 4, as we argued in Case 1.

Case 3. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple affine algebra $\mathbf{A}$ with a 1-element subuniverse (so $H = \mathrm{GL}(n, q) \ltimes T$).

As in Case 2, it suffices to prove that $n > 1$ cannot occur. Suppose otherwise that $H = \mathrm{GL}(n, q) \ltimes T$ and $n > 1$. If q is odd, then H contains at least three mutually commuting involutions of the form $L\mathbf{x} + K$ where $K = 0$ and L is a diagonal matrix with ± 1's distributed along the diagonal in different ways. If q is even and the parameter d from Theorem 5.9 is nonzero, then the group of translations is an elementary abelian 2-group of size q^{nd}, so there exist distinct commuting involutions.

Thus $d = 0$, and $H = \mathrm{GL}(n, q)$ for some $n > 1$ and some $q = 2^k$. Consider involutions of the form $L_a\mathbf{x}$ and $L'\mathbf{x}$ where

$$L_a = \begin{bmatrix} 1 & a & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{with } a \in \mathbb{F}^\times, \text{ and} \quad L' = \begin{bmatrix} 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

If $q > 2$, then there exist $a \neq b$ in $\mathbb{F}_q^\times$, and $L_a\mathbf{x}$ and $L_b\mathbf{x}$ are distinct commuting involutions in H. If $n > 2$, then $L_1\mathbf{x}$ and $L'\mathbf{x}$ are distinct commuting involutions in H. Thus $q = n = 2$, and $H = \mathrm{GL}(2, 2)$ has order 6. Now the 2-Sylow subgroup of H is not big enough to be generalized quaternion or cyclic of order at least 4. This completes the proof. □

Corollary 6.7. *If G has a 2-Sylow subgroup that is a generalized quaternion group of more than eight elements, then G is not the automorphism group of a finite, 2-generated, free algebra.*

Proof. Suppose that G has a 2-Sylow subgroup that is a generalized quaternion group. By Theorem 6.6, there is a surjective homomorphism $\varphi\colon G \twoheadrightarrow H$ where H is sharply 2-transitive and has an isomorphic 2-Sylow subgroup. By Theorem 5.2, if a sharply 2-transitive group has a 2-Sylow subgroup that is generalized quaternion, then the 2-Sylow subgroup is the 8-element quaternion group. □

Remark 6.8. In fact, this proof shows that even the 8-element quaternion group rarely occurs as the 2-Sylow subgroup of the automorphism group G of a finite, 2-generated free algebra. It can occur only if G has a homomorphism onto some exceptional 2-transitive group of degree $5^2, 11^2, 11^2, 29^2$, or 59^2.

Conversely, if there is an *isomorphism* of G onto any sharply 2-transitive group, then G may be realized as the automorphism group of a finite, 2-generated free algebra. For suppose that G acts sharply 2-transitively on A. Let $\mathbf{A}$ be the algebra with universe A and operations consisting of all idempotent operations on A that commute with the permutations in G. Then $\mathbf{A}$ is a strictly simple quasiprimal algebra. If $\mathbf{E}$ is the 2-generated free algebra in the variety generated by $\mathbf{A}$, then $\mathrm{Aut}(\mathbf{E}) \cong G$.

Theorem 6.9. *If $n > 1$ and q is an odd prime power, then $\mathrm{SL}(n, q)$ is not the automorphism group of a finite, 2-generated, free algebra.*

Proof. We consider the case where $(n, q) = (2, 3)$ first. The matrix $-I \in Z(\mathrm{SL}(2, 3))$ is the unique involution of $\mathrm{SL}(2, 3)$, and $\mathrm{SL}(2, 3)$ has no retraction onto $\{\pm I\}$. By Corollary 6.4, $\mathrm{SL}(2, 3)$ is not representable as the automorphism group of a finite, 2-generated, free algebra.

When $n > 1$, q is an odd prime power, and $(n, q) \neq (2, 3)$, the group $G = \mathrm{SL}(n, q)$ is quasisimple. This property is preserved by nonconstant surjective homomorphisms, so if $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ then H_σ is quasisimple.

Case 1. $H = S_n$ and σ is a fixed point free involution.

S_n is not quasisimple, so this case does not arise.

Case 2. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple quasiprimal algebra $\mathbf{A}$ (so $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$).

$\mathrm{Aut}(\mathbf{A}) \wr S_n$ is not quasisimple when $n > 1$, since it has a homomorphism onto S_n. Thus, $n = 1$, $\mathbf{A}$ is strictly simple and free, and H acts sharply 2-transitively on A. The Frobenius kernel K of H satisfies $C_H(K) = K$. Since the proper normal subgroups of H lie in the center, $H = C_H(K) = K$, contradicting the fact that H is quasisimple.

Case 3. H_σ is the automorphism group of a free algebra in a variety generated by a strictly simple affine algebra $\mathbf{A}$ with a 1-element subuniverse (so $H = \mathrm{GL}(k, r) \ltimes T$).

$H = \mathrm{GL}(k, r) \ltimes T$ is quasisimple if and only if T is trivial, $k > 1$, and $r = 2$. It follows from Theorem 5.9 that if T is trivial, then parameter $d = 0$, so $k = 2m$ is even. Thus, if H is quasisimple, then $H = \mathrm{GL}(2m, 2) = \mathrm{PSL}(2m, 2)$ is simple. A surjective homomorphism $\varphi\colon G_\sigma \to H_\sigma$ induces an isomorphism between their simple factors, i.e., $\mathrm{PSL}(n, q) \cong \mathrm{PSL}(2m, 2)$. But it follows from the results of [1] that if $\mathrm{PSL}(n, q) \cong \mathrm{PSL}(2m, 2)$, then $(n, q) = (2m, 2)$. Since q is odd, this does not happen. □

Theorem 6.10. *Let G be a simple group with designated involution σ. If G_σ forces congruence permutability, then G_σ is the automorphism group of a finite, 2-generated, free algebra if and only if $G_\sigma \cong \mathrm{GL}(2m, 2)_\sigma$ for some $m \geq 2$, where $\sigma \in \mathrm{GL}(2m, 2)$ is the matrix*

$$\begin{bmatrix} S & 0 & \cdots & 0 \\ 0 & S & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & S \end{bmatrix}, \qquad \text{and } S \text{ is the matrix } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

This involution is not 2-central.

Proof. For the "if" part of the theorem, $\mathrm{GL}(2m, 2) = \mathrm{PSL}(2m, 2)$ is simple when $m \geq 2$. By Theorem 5.9, $\mathrm{GL}(2m, 2)_\sigma$ arises as the automorphism group of the 2-generated free algebra in a minimal congruence permutable variety generated by a strictly simple affine algebra with exactly one 1-element subuniverse. More explicitly, it is the automorphism group of a 2-generated free module over the ring of $2m \times 2m$-matrices over $\mathbb{F}_2$.

Conversely, if G_σ forces congruence permutability, then there must be a homomorphism $\varphi\colon G_\sigma \twoheadrightarrow H_\sigma$ onto one of our special groups H_σ. This homomorphism is nonconstant, since it preserves σ. G is simple, so $G_\sigma \cong H_\sigma$. But the only simple groups that can occur as the target group H_σ in Theorems 4.13, 5.5 and 5.9 are the groups of the form $\mathrm{GL}(2m, 2)$, $m \geq 2$, from Theorem 5.9. This group arises only when the parameter d equals 0, and when that happens the matrix for σ is the one described in the theorem statement.

Now we argue that this involution is not 2-central. The group U of unipotent matrices (i.e., matrices of the form $I + N$ where N is strictly upper triangular) is a 2-Sylow subgroup of $\mathrm{GL}(2m, 2)$. Let $E_{i,j}$ be the matrix with 1 in the ij-th position and 0's elsewhere. The only nonidentity matrix $I + N \in U$ that commutes with all $I + E_{i,j}$, $i < j$, is $I + E_{1,2m}$, so U has a unique central involution. Therefore every 2-Sylow subgroup of $\mathrm{GL}(2m, 2)$ has a unique central involution, implying that the set of 2-central involutions equals the conjugacy class of $I + E_{1,2m}$. But the matrix for σ from Theorem 6.10 is not in the conjugacy class of $I + E_{1,2m}$ when $m \geq 2$, since the number of Jordan blocks for σ is m and for $I + E_{1,2m}$ is $2m - 1$. □

Corollary 6.11. *Let $\sigma \in G$ be a 2-central involution. If G is a (non-sporadic) Suzuki simple group, a Mathieu simple group, or A_n with $n \equiv 0, 1 \pmod 4$ and $n \geq 8$, then G_σ is not the automorphism group of a finite, 2-generated, free algebra.*

Proof. If G is one of these groups and σ is a 2-central involution, then it follows from Section 3 that G_σ forces congruence permutability. By Theorem 6.10, these groups are not representable. □

Remark 6.12. Since $A_8 \cong \mathrm{PSL}(4, 2) = \mathrm{GL}(4, 2)$ (cf. [1]), A_8 is representable as $\mathrm{Aut}(\mathbf{F})$ when $\mathbf{F} = \mathbf{F}_\mathcal{V}(x, y)$ is the 2-generated free module over the ring of 2×2-matrices over $\mathbb{F}_2$. By Corollary 6.11, $(A_8)_\sigma$ is not representable as $\mathrm{Aut}(\mathbf{F})_\sigma$ when σ is a 2-central involution. Thus $(A_8)_\sigma$ illustrates that a simple group may be representable when σ is from one conjugacy classes of involutions and not representable when σ is from another class. Since A_8 has only two conjugacy classes of involutions, we have complete information concerning the representability of this group.

Gould's constructions in [4] of algebras whose squares have prescribed automorphism groups have an extra property that proves the following when translated into the the free algebra setting:

(†) If G is any group with designated involution σ, then G_σ is the automorphism group of a 2-generated, *idempotent*, free algebra.

(‡) If G is a finite group with designated involution σ and there is a retraction of G onto $\langle\sigma\rangle$, then G_σ is the automorphism group of a finite, 2-generated, *idempotent*, free algebra. (Cf. Theorem 2.16 (1).)

This raises the question of whether or not every group representable as $\mathrm{Aut}(\mathbf{F})_\sigma$, where $\mathbf{F} = \mathbf{F}_\mathcal{V}(x, y)$ is finite, is representable when $\mathbf{F}$ is finite and idempotent. If

so, then according to the proof of Theorem 2.3 it could be arranged that $\mathbf{F}$ lies in $\mathcal{V}[G_\sigma, G]$. When G has a retraction onto $\langle\sigma\rangle$, then there is such an $\mathbf{F}$, by (‡). When G does not have a retraction onto $\langle\sigma\rangle$, then from Theorem 2.10 we know at least that $\mathcal{V}[G_\sigma, G]$ is congruence permutable, so the techniques of this paper may be applied.

Theorem 6.13. *Let G is a finite group with designated involution σ. If G has no retraction onto $\langle\sigma\rangle$ and G_σ is representable as $Aut(\mathbf{F}_\mathcal{V}(x,y))_\sigma$ for some finite free algebra in an idempotent variety, then there is a surjective homomorphism $\varphi\colon G_\sigma \to H_\sigma$ where*

1. *$H = S_n$, σ is a fixed point free involution, and $n \equiv 0 \pmod 4$.*
2. *H_σ is as in Theorem 5.5 with k even and $\ell = 0$.*
3. *H_σ is as in Theorem 5.9 with $d = m$. (In particular, $H = \mathrm{GL}(n,q) \ltimes T$ where $|T| = q^{nd} = q^{n^2} > 1$.)*

Proof. If G has no retraction onto $\langle\sigma\rangle$, then G_σ forces congruence permutability for idempotent varieties. There must be a surjective homomorphism $\varphi\colon G_\sigma \to H_\sigma$ where H_σ is as in Theorem 4.13, 5.5 or 5.9.

If H_σ is as in Theorem 4.13, then $H = S_n$ and σ is a fixed point free involution. The existence of a fixed point free involution forces $n \equiv 0 \pmod 2$, but if $n \not\equiv 0 \pmod 4$ then σ is an odd permutation. In this case H has a retraction onto $\langle\sigma\rangle$, so composing φ with that retraction leads to a retraction of G onto $\langle\sigma\rangle$. Thus it must be that $n \equiv 0 \pmod 4$.

If H_σ is as in Theorem 5.5, then $H = \mathrm{Aut}(\mathbf{A}) \wr S_n$. Here, since $\mathbf{A}$ is idempotent, every diagonal pair of $A \times A$ lies in a bad orbit, so there are no diagonal nonisolated pairs. This implies that $\ell = 0$. If there are an odd number k of isolated pairs, then the surjective homomorphism $\psi\colon \mathrm{Aut}(\mathbf{A}) \wr S_n \to S_n\colon (\alpha, \beta) \mapsto \beta$ maps σ to an odd permutation. Composing φ with ψ and then a retraction onto $\langle\sigma\rangle$ leads to a retraction of G onto $\langle\sigma\rangle$.

If H_σ is as in Theorem 5.9, then the parameter m is the $\mathbb{F}_q$-dimension of $\mathbf{A}$ while the parameter d is the $\mathbb{F}_q$-dimension of the space of 1-element subuniverses of $\mathbf{A}$. If $\mathbf{A}$ is idempotent, then $d = m = 2m - d = n$. □

Corollary 6.14. *If G is a finite, nonabelian, simple group with designated involution σ, then G_σ is not representable as the automorphism group of a finite, 2-generated,* idempotent*, free algebra.*

Proof. The simplicity of G forces $G_\sigma \cong H_\sigma$ for one of the pointed groups H_σ of Theorem 6.13. But none of these groups is simple. □

Remark 6.15. As noted in Remark 6.12, the alternating group A_8 is representable as the automorphism group of a finite, 2-generated, free module. Corollary 6.14 shows that is not representable as the automorphism group of any finite, idempotent, 2-generated, free algebra.

References

[1] E. Artin, *The orders of the linear groups*, Comm. Pure Appl. Math. **8** (1955), 355–365.
[2] R. Freese and R. McKenzie, *Commutator theory for congruence modular varieties*, London Mathematical Society Lecture Note Series, **125**. Cambridge University Press, Cambridge, 1987.
[3] O. C. García and W. Taylor, *The lattice of interpretability types of varieties*, Mem. Amer. Math. Soc. **50** (1984), no. 305.
[4] M. Gould, *Endomorphism and automorphism structure of direct squares of universal algebras*, Pacific J. Math. **59** (1975), no. 1, 69–84.
[5] M. Gould, *Automorphism groups of free algebras and direct powers*, Universal algebra (Esztergom, 1977), pp. 331–334, Colloq. Math. Soc. János Bolyai, **29**, North-Holland, Amsterdam-New York, 1982.
[6] D. Gorenstein, R. Lyons, R. Solomon, *The Classification of Finite Simple Groups*, Mathematical Surveys and Monographs **40**, no. 1, American Mathematical Society, 1994.
[7] B. Huppert and N. Blackburn, *Finite Groups, III*, Grundlehren der mathematischen Wissenschaften **243**, Springer-Verlag, 1982.
[8] K. A. Kearnes and Á. Szendrei, *A characterization of minimal locally finite varieties*, Trans. Amer. Math. Soc. **349** (1997), no. 5, 1749–1768.
[9] Á. Szendrei, *Clones in universal algebra*, Séminaire de Mathématiques Supérieures [Seminar on Higher Mathematics], **99**. Presses de l'Université de Montréal, Montreal, QC, 1986.
[10] S. T. Tschantz *A simple variety with a long Maltsev term: solution of the* $\mathbb{Z}_4$ *problem*, unpublished manuscript, 1996.

(Keith Kearnes) Department of Mathematics, University of Colorado, Boulder CO 80309, USA
E-mail address: kearnes@euclid.colorado.edu

(Steven Tschantz) Department of Mathematics, Vanderbilt University, Nashville TN 37240, USA
E-mail address: tschantz@math.vanderbilt.edu