Supplementary Materials

Here, we present the grid search results for the gravity compensated and uncompensated system with ACKTR (Table S1 and S2) and PPO2 (Table S3 and S4). Each hyperparameter permutation was used in 10 training sessions. For each 10 training sessions, we compute the mean and standard deviation of cumulative regret (μ_{regret} and σ_{regret}, respectively), as well as the mean and standard deviation of the final episodic reward (μ_{FER} and σ_{FER}, respectively). We rank each set of hyperparameters in order of ascending mean cumulative regret. For brevity, we only include the hyperparameters of each algorithm that took on multiple values in the grid search.

The grid search results for the gravity uncompensated system with ACKTR and PPO2 are shown in Table S1 and S3, respectively, and gravity compensated system with ACKTR and PPO2 are listed in Table S2 and S4, respectively. *ent_coef* shows the coefficient of the entropy loss. In the Stable Baselines implementation of ACKTR, the mean entropy of the output distribution of both the actor and the critic is penalized by a factor of *ent_coef.* We abbreviate the learning rate as *lr*, and *lr_sch* represents a way to change the learning rate during the training; *linear* means that the learning rate is linearly decreased, *constant* means the learning rate is fixed to one value, and double middle drop means that the learning rate is decreased nonlinearly. We abbreviate the double middle drop as *dmd. vf_coef* is the coefficient of the value function loss. In ACKTR, the actor and critic share the same neural network layers and are separate only in the output layer. Thus, the loss of the critic and actor are summed together to train the network. The value of *vf_coef* scales the loss of the critic in this summation. The following shows which values are used for the rest of parameters in the two agents: gamma = 0.99, n_steps = 32, vf_fisher_coef = 1, max_grid_norm = 0.5, and kfac_clip = 0.001 for ACKTR and gamma = 0.99, max_grad_norm = 0.5, lam = 0.95, nuinibathes = 4, noptepoches = 4, cliprange = 0.2, and cliprange_vf = 0.2 for PPO2. The definition of each parameter can be found in [21].

In addition, to evaluate the system's behavior while acting under the agents' policy, we generated histograms of the states visited and actions taken over the course of 10 episodes. We selected one agent from the 100 trained for each experiment. We selected agents by iterating over the 100 and finding the first trained agent that was able to obtain more than 5.5×10^5 episodic reward. It is acknowledged that different agents trained in the same experiment may exhibit different behavior, but we found the general behavior of agents operating on the same task to be very similar. Figures S1–S20 are histograms generated based on the state and action data.

Table S1. ACKTR uncompensated system grid search results.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
1	0	0.12	linear	0.75	44,063	13,486	23,035	8997
2	0.001	0.12	linear	1	44,434	13,875	22,879	9182
3	0.001	0.12	constant	1	44,444	13,733	21,748	14,817
4	0	0.12	linear	1	44,493	13,545	23,081	12,641
5	0	0.12	constant	1	44,900	14,225	19,372	13,969
6	0.001	0.3	constant	1	44,987	12,778	23,474	17,867
7	0.01	0.12	linear	0.75	45,408	12,404	27,909	8388
8	0	0.12	constant	0.75	45,838	12,852	23,350	14,437
9	0.01	0.12	constant	0.5	45,997	12,316	27,673	5946
10	0.001	0.02	constant	0.75	46,024	12,566	26,595	12,952
11	0.01	0.02	constant	0.75	46,097	12,184	26,615	10,107
12	0	0.05	linear	1	46,165	12,431	30,296	6380
13	0.001	0.12	constant	0.5	46,235	13,039	22,267	13,223
14	0	0.05	linear	0.5	46,268	12,438	30,423	6213
15	0.01	0.05	constant	0.75	46,297	12,403	25,806	9638
16	0.001	0.05	constant	1	46,395	12,578	24,874	11,788
17	0	0.02	constant	1	46,480	12,425	25,989	11,097
18	0.001	0.05	linear	0.5	46,538	11,240	31,204	7785
19	0	0.05	linear	0.75	46,615	11,518	32,187	7440
20	0.001	0.05	constant	0.75	46,666	12,549	28,255	14,967
21	0.001	0.05	constant	0.5	46,743	12,346	24,790	14,257
22	0.001	0.05	linear	1	46,792	11,841	31,361	10,204
23	0.001	0.12	linear	0.75	46,813	12,232	28,738	12,887
24	0.001	0.12	linear	0.5	47,012	11,863	28,894	11,397
25	0	0.05	constant	1	47047	12,398	25,937	11,398
26	0.01	0.02	constant	1	47056	11,907	28,163	6894
27	0	0.05	constant	0.5	47075	11,765	27,745	11,778
28	0	0.02	constant	0.5	47219	11,228	27,763	14,758

(Table S1 continued on next page)

Table S1. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
29	0	0.3	constant	0.75	47,258	12,551	24,962	9952
30	0.001	0.02	constant	0.5	47,415	11,415	29,213	12,600
31	0.001	0.3	linear	0.75	47,438	10,931	29,872	19,197
32	0	0.02	constant	0.75	47,496	11,928	26,936	13,688
33	0.01	0.05	linear	1	47,593	11,173	34,640	9625
34	0	0.3	dmd	1	47,857	12,264	27,244	13,130
35	0.01	0.12	linear	0.5	47,916	10,734	32,521	11,922
36	0.001	0.02	constant	0.25	48,189	11,257	27,407	8871
37	0.01	0.3	dmd	0.75	48,331	9905	32,950	14,448
38	0.01	0.05	constant	1	48,336	11,246	30,009	13,025
39	0.001	0.05	linear	0.75	48,495	11,217	33,390	10,293
40	0	0.05	constant	0.75	48,541	11,422	29,295	9849
41	0.001	0.3	constant	0.75	48,552	10,280	30,042	15,425
42	0.01	0.05	linear	0.75	48,628	10,247	35,564	9641
43	0.01	0.12	constant	0.75	48,670	10,949	30,418	14,644
44	0.01	0.12	constant	1	48,683	10,538	32,044	10,902
45	0.01	0.05	constant	0.5	48,713	10,711	31,430	11,285
46	0.001	0.3	linear	1	48,773	11,518	29,262	8081
47	0.001	0.3	dmd	1	48,817	10,413	33,353	15,704
48	0.01	0.02	constant	0.25	48,981	10,472	32,518	7247
49	0	0.05	linear	0.25	49,468	9872	36,558	11,513
50	0	0.02	linear	1	49,515	9555	37,655	4837
51	0	0.3	constant	0.5	49,546	10,234	31,091	14,381
52	0.001	0.02	linear	0.5	49,692	8926	38,477	5190
53	0.001	0.3	constant	0.25	49,838	9483	34,642	17,282
54	0.01	0.3	constant	0.75	49,951	9692	34,944	15,921
55	0.01	0.05	linear	0.5	50,189	9263	37,616	9693
56	0	0.02	constant	0.25	50,288	10,173	33,641	12,191
57	0	0.12	constant	0.5	50,291	9517	35,051	13,793
58	0.01	0.3	constant	1	50,319	9417	33,799	9895
59	0	0.3	linear	0.75	50,478	9394	33,630	16,003
60	0	0.12	dmd	1	50,666	10,056	35,018	5756
61	0.001	0.12	constant	0.75	50,976	9362	35,081	12,478
62	0.01	0.12	dmd	0.5	51,006	8505	36,920	7156
63	0.001	0.12	linear	0.25	51,013	8891	36,769	14,309
64	0.01	0.12	linear	1	51,034	9730	35,731	8184
65	0.01	0.12	dmd	1	51,092	8492	35,727	4699
66	0.001	0.3	linear	0.5	51,108	9342	35,579	12,941
67	0.01	0.12	dmd	0.75	51,133	8594	36,038	7638

(Table S1 continued on next page)

Table S1. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
68	0.01	0.3	linear	0.5	51,349	8576	38,316	9066
69	0.01	0.3	linear	1	51,450	9152	37,052	9942
70	0	0.02	linear	0.75	51,518	8171	42,084	7216
71	0.001	0.05	constant	0.25	51,597	8678	37,964	18,484
72	0	0.12	linear	0.5	51,707	8461	37,968	14,587
73	0	0.12	linear	0.25	51,725	8563	38,505	11,413
74	0.001	0.3	dmd	0.75	51,733	8560	36,946	14,285
75	0.001	0.02	constant	1	51,736	9092	36,864	15,827
76	0	0.05	constant	0.25	51,820	8185	38,079	18,767
77	0.001	0.02	linear	0.25	51,838	7485	42,482	7872
78	0	0.12	constant	0.25	51,882	8455	36,840	16,589
79	0.01	0.12	linear	0.25	52,102	8814	38,398	10,906
80	0	0.12	dmd	0.75	52,317	8221	38,526	7551
81	0.001	0.12	dmd	0.75	52,378	7710	37,677	8186
82	0.01	0.02	linear	1	52,498	7536	44,647	7363
83	0.01	0.3	constant	0.25	52,503	7413	40,492	12,478
84	0.001	0.02	linear	0.75	52,566	7133	44,159	7268
85	0.001	0.12	dmd	1	52,570	8574	36,088	6505
86	0.01	0.05	linear	0.25	52,571	7215	42,519	9780
87	0.01	0.02	linear	0.75	52,581	7655	44,017	6306
88	0	0.3	constant	0.25	52,598	8144	38,784	17,335
89	0	0.3	dmd	0.75	52,613	8222	38,942	16,020
90	0.001	0.12	dmd	0.5	52,623	7242	39,983	9402
91	0.01	0.3	dmd	1	52,757	8210	39,427	9389
92	0.001	0.02	linear	1	52,834	7665	43,052	6775
93	0	0.12	dmd	0.5	52,879	7094	41,169	9707
94	0.01	0.02	constant	0.5	52,897	7264	41,797	13,674
95	0.01	0.3	constant	0.5	52,911	7890	40,208	14,396
96	0	0.3	constant	1	53,207	7537	40,725	13,317
97	0.001	0.05	linear	0.25	53,331	7007	44,643	11,044
98	0.001	0.3	constant	0.5	53,354	7204	40,953	16,577
99	0.001	0.12	constant	0.25	53,452	7812	40,240	12,819
100	0.01	0.05	constant	0.25	53,680	7140	41,127	10,293
101	0.01	0.02	linear	0.5	53,724	6277	45,908	9548
102	0	0.12	dmd	0.25	54,064	6252	42,353	11,560
103	0	0.02	linear	0.25	54,165	6033	47,736	7576
104	0	0.02	linear	0.5	54,186	6509	46,854	10,483
105	0	0.3	dmd	0.5	54,191	6595	43,988	13,541
106	0.001	0.3	linear	0.25	54,233	6326	43,811	10,244

(Table S1 continued on next page)

Table S1. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
107	0.01	0.3	linear	0.75	54,413	6417	44,629	12,068
108	0.1	0.12	constant	1	54,618	5590	49,687	2924
109	0.001	0.12	dmd	0.25	54,668	7439	40,884	10,259
110	0	0.3	linear	1	54,707	7137	43,037	14,786
111	0.1	0.12	linear	1	54,755	5940	48,728	2782
112	0.1	0.02	constant	0.75	54,921	5428	48,877	3059
113	0.1	0.02	constant	0.5	55,114	5046	51,046	3574
114	0.1	0.12	constant	0.5	55,121	5205	51,695	2826
115	0.01	0.12	dmd	0.25	55,159	6299	43,548	7354
116	0.1	0.05	linear	0.5	55,241	5145	50,063	4600
117	0.01	0.02	linear	0.25	55,315	5174	48,637	5544
118	0.1	0.02	constant	1	55,324	5503	48,524	4983
119	0.1	0.05	linear	0.75	55,813	4960	50,550	3090
120	0.01	0.12	constant	0.25	55,829	5454	47,480	10,050
121	0.1	0.12	linear	0.5	55,852	4807	51,090	3716
122	0.001	0.3	dmd	0.5	55,854	6331	44,784	13,517
123	0.1	0.05	linear	1	55,856	4899	51,007	4904
124	0	0.3	dmd	0.25	55,863	5367	45,410	14,595
125	0	0.05	dmd	0.75	55,914	4746	48,092	4978
126	0.1	0.05	constant	0.75	55,950	4960	51,978	2552
127	0.001	0.05	dmd	0.75	56,015	4210	49,571	4827
128	0.1	0.12	linear	0.75	56,097	4790	50,710	3887
129	0.1	0.3	constant	0.75	56,168	5016	51,546	4356
130	0.1	0.02	linear	0.75	56,178	4861	50,531	3936
131	0.1	0.3	linear	0.5	56,233	4576	51,487	2389
132	0.1	0.02	linear	1	56,334	4784	50,832	4514
133	0.1	0.12	dmd	0.75	56,340	3990	51,978	2101
134	0.1	0.02	linear	0.5	56,401	4441	52,139	2645
135	0.1	0.12	constant	0.75	56,517	4273	52,461	3952
136	0.1	0.05	constant	0.25	56,531	3948	53,708	2223
137	0.1	0.3	constant	1	56,574	4869	51,069	4619
138	0.1	0.02	linear	0.25	56,678	4822	51,492	4274
139	0.01	0.05	dmd	1	56,699	4140	49,510	3343
140	0.01	0.3	linear	0.25	56,728	5126	48,694	11,632
141	0	0.3	linear	0.5	56,738	5298	47,951	13,372
142	0.01	0.3	dmd	0.25	56,809	5953	46,473	12,847
143	0.1	0.3	dmd	1	56,908	4913	51,184	3808
144	0.1	0.3	linear	1	56,913	4411	51,231	3499
145	0.1	0.05	constant	0.5	56,927	3888	53,503	4112

(Table S1 continued on next page)

Table S1. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
146	0	0.02	dmd	0.5	56,939	3457	52,350	4722
147	0.1	0.05	constant	1	56,953	4386	51,692	4175
148	0.1	0.3	linear	0.75	56,961	4431	53,115	3095
149	0.1	0.02	constant	0.25	57,054	4596	52,351	4395
150	0.1	0.05	linear	0.25	57,158	4124	52,827	3904
151	0.01	0.05	dmd	0.75	57,160	4052	50,264	4659
152	0.01	0.3	dmd	0.5	57,276	5027	48,543	8948
153	0	0.05	dmd	1	57,308	4038	49,644	4906
154	0.001	0.3	dmd	0.25	57,462	5203	47,818	14,296
155	0.001	0.05	dmd	0.5	57,492	3719	51,119	4061
156	0	0.05	dmd	0.5	57,646	3543	51,240	5004
157	0.1	0.12	constant	0.25	57,648	3284	55,184	3599
158	0.1	0.3	dmd	0.75	57,664	3433	55,340	4024
159	0.1	0.12	dmd	1	57,671	3743	52,673	3089
160	0.001	0.05	dmd	1	57,685	3580	52,157	5107
161	0	0.3	linear	0.25	57,686	5669	47,860	13,218
162	0.01	0.05	dmd	0.5	57,785	3600	51,798	3319
163	0.1	0.12	dmd	0.5	57,887	3846	52,792	3984
164	0.1	0.3	dmd	0.5	58,027	3648	54,896	4859
165	0	0.02	dmd	0.75	58,199	2751	55,087	2355
166	0.001	0.02	dmd	0.5	58,254	2751	55,109	4514
167	0.01	0.05	dmd	0.25	58,308	3115	53,176	4873
168	0.1	0.12	linear	0.25	58,389	3504	54,440	3507
169	0.001	0.02	dmd	1	58,410	2871	54,365	5028
170	0.01	0.02	dmd	0.5	58,478	2737	55,195	4422
171	0	0.05	dmd	0.25	58,601	3164	53,556	4199
172	0.1	0.05	dmd	0.75	58,662	2867	55,451	2844
173	0.1	0.05	dmd	0.5	58,698	2841	54,955	2108
174	0.1	0.3	constant	0.5	58,705	3167	55,401	3213
175	0.1	0.12	dmd	0.25	58,767	3079	54,983	3388
176	0.001	0.05	dmd	0.25	58,829	3446	51,665	6962
177	0	0.02	dmd	1	58,831	2503	56,314	2596
178	0.1	0.3	constant	0.25	59,084	3300	56,982	1842
179	0.1	0.05	dmd	0.25	59,127	2458	55,681	2970
180	0.1	0.3	linear	0.25	59,303	3052	55,709	3754
181	0.1	0.05	dmd	1	59,353	2728	55,394	2554
182	0.01	0.02	dmd	1	59,624	2247	57,300	1748
183	0.01	0.02	dmd	0.25	59,762	2308	56,420	3694
184	0.001	0.02	dmd	0.25	59,826	2063	57,842	3422

(Table S1 continued on next page)

Table S1. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
185	0.1	0.3	dmd	0.25	59,829	2780	56,892	4242
186	0.001	0.02	dmd	0.75	59,862	2262	56,919	4729
187	0.01	0.02	dmd	0.75	59,913	2152	57,322	1965
188	0.1	0.02	dmd	0.25	60,323	1755	58,620	3040
189	0.1	0.02	dmd	1	60,450	1884	58,268	1983
190	0	0.02	dmd	0.25	60,766	1878	57,723	3170
191	0.1	0.02	dmd	0.5	60,861	1723	59,250	1720
192	0.1	0.02	dmd	0.75	61,181	1499	59,346	2859

Table S2. ACKTR gravity compensated system grid search results.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
1	0.001	0.05	constant	1	37,132	18,116	10,459	7682
2	0	0.12	linear	0.75	37,210	18,639	12,170	9293
3	0.01	0.12	constant	1	37,802	17,778	12,695	2220
4	0.001	0.05	linear	0.5	38,039	17,442	17,331	8394
5	0	0.12	constant	0.75	38,137	17,580	13,683	14,598
6	0.001	0.02	constant	1	38,240	18,308	10,910	6462
7	0	0.05	constant	0.75	38,298	17,098	14,280	13,416
8	0	0.02	constant	0.75	38,359	17,843	11,828	8542
9	0.01	0.05	constant	1	38,363	17,784	14,133	7517
10	0.001	0.05	constant	0.75	38,565	17,712	13,143	11,368
11	0.001	0.02	constant	0.25	38,736	16,819	14,809	14,936
12	0.001	0.05	linear	1	38,825	16,629	18,117	10,601
13	0.001	0.12	linear	0.75	39,037	17,537	14,475	8640
14	0.001	0.12	linear	0.25	39,076	17,048	14,805	9769
15	0.001	0.12	constant	0.5	39,296	17,263	12,773	10,209
16	0.001	0.05	linear	0.75	39,339	17,077	17,247	8288
17	0.01	0.05	linear	1	39,469	16,425	19,161	4252
18	0.01	0.02	constant	0.5	39,473	15,553	21,217	11,829
19	0.01	0.02	constant	0.75	39,820	16,438	15,702	9417
20	0.001	0.02	constant	0.5	40,000	15,156	20,470	21,363
21	0	0.05	linear	1	40,103	16,330	20,560	14,670
22	0.01	0.02	constant	1	40,169	16,752	14,888	5677
23	0	0.02	constant	0.25	40,444	15,897	17,116	17,296
24	0.01	0.05	constant	0.25	40,496	17,125	15,716	8790
25	0.01	0.3	constant	0.75	40,580	15,986	16,766	11,270
26	0	0.05	constant	0.5	40,598	15,699	16,797	16,731

(Table S2 continued on next page)

Table S2. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
27	0.001	0.3	constant	0.75	40,894	15,199	19,464	19,525
28	0	0.3	constant	0.75	40,934	14,888	18,793	16,270
29	0.001	0.05	constant	0.25	40,942	15,802	17,726	18,101
30	0	0.05	constant	1	40,972	16,573	16,127	15,455
31	0.001	0.12	constant	1	41,013	15,887	17,693	14,488
32	0	0.05	constant	0.25	41,269	15,216	19,946	18,676
33	0	0.12	constant	1	41,274	16,320	15,876	16,667
34	0.01	0.12	linear	0.75	41,444	15,796	20,013	11,996
35	0.01	0.12	constant	0.25	41,500	14,636	18,916	15,065
36	0	0.05	linear	0.75	41,542	15,497	22,283	14,664
37	0.01	0.12	constant	0.75	41,551	15,077	18,939	9994
38	0	0.3	linear	0.75	42,222	14,646	21,192	19,597
39	0.001	0.05	linear	0.25	42,236	14,137	24,305	15,012
40	0.001	0.3	linear	1	42,236	15,758	21,912	16,286
41	0.001	0.3	linear	0.5	42,281	14,788	21,558	17,239
42	0	0.02	constant	0.5	42,410	15,146	18,620	16,555
43	0	0.12	linear	1	42,432	14,980	20,706	13,033
44	0	0.12	linear	0.5	42,456	15,098	20,080	13,895
45	0	0.05	linear	0.25	42,469	14,174	24,099	14,039
46	0.01	0.05	linear	0.75	42,596	14,257	25,654	13,621
47	0.001	0.3	constant	1	42,704	13,644	23,933	20,847
48	0	0.12	constant	0.5	42,780	15,135	18,581	10,835
49	0.001	0.12	linear	0.5	42,876	14,027	22,500	19,167
50	0.01	0.12	linear	0.5	43,325	13,484	24,096	12,983
51	0.01	0.02	linear	0.75	43,342	12,914	29,076	6423
52	0.001	0.02	constant	0.75	43,350	14,142	23,322	17,960
53	0.01	0.05	constant	0.5	43,581	13,921	24,219	14,468
54	0.01	0.05	constant	0.75	43,705	14,017	23,254	16,542
55	0.001	0.3	dmd	0.5	43,734	13,510	24,625	22,243
56	0.01	0.12	linear	1	43,894	13,872	24,519	17,668
57	0.001	0.12	linear	1	43,984	14,355	22,591	15,422
58	0.01	0.3	dmd	0.5	44,023	14,647	21,287	16,815
59	0.001	0.05	constant	0.5	44,104	13,779	22,938	16,746
60	0.001	0.3	dmd	1	44,407	14,106	23,486	20,143
61	0.001	0.02	linear	1	44,552	12,364	30,691	12,248
62	0	0.02	linear	0.75	44,644	12,031	31,985	13,175
63	0	0.3	linear	1	44,762	13,214	24,721	17,072
64	0.01	0.3	linear	0.75	45,049	13,386	24,196	17,627
65	0.01	0.02	constant	0.25	45,050	12,463	27,551	16,470

(Table S2 continued on next page)

Table S2. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
66	0.001	0.12	constant	0.25	45,072	12,656	25,076	16,858
67	0.01	0.05	linear	0.5	45,229	12,174	29,899	17,567
68	0	0.12	constant	0.25	45,250	13,439	23,224	14,948
69	0.01	0.12	constant	0.5	45,720	12,943	25,572	17,041
70	0.001	0.12	dmd	0.75	45,838	13,261	23,440	10,273
71	0	0.3	constant	1	45,845	14,047	22,520	11,336
72	0.01	0.05	linear	0.25	46,228	12,541	29,220	13,902
73	0.01	0.02	linear	1	46,306	11,550	32,417	10,584
74	0	0.02	linear	0.5	46,306	10,968	32,790	9056
75	0.001	0.12	constant	0.75	46,445	12,126	26,862	19,419
76	0	0.05	linear	0.5	46,676	11,548	32,041	19,113
77	0.01	0.12	dmd	0.75	46,764	11,935	26,183	9856
78	0.01	0.3	linear	1	46,940	11,754	27,854	12,014
79	0.01	0.02	linear	0.5	46,945	10,695	34,185	10,360
80	0	0.12	dmd	0.75	47,010	11,248	28,702	12,912
81	0.001	0.02	linear	0.75	47,065	10,495	34,263	12,279
82	0	0.3	dmd	0.75	47,282	11,601	30,251	22,998
83	0	0.3	constant	0.25	47,596	11,687	29,005	18,877
84	0	0.02	linear	1	47,609	11,149	33,765	7627
85	0	0.3	dmd	0.25	47,644	10,781	32,622	21,887
86	0.01	0.12	dmd	1	47,656	11,596	26,897	7371
87	0.001	0.12	dmd	1	47,766	11,170	26,782	10,622
88	0	0.3	dmd	0.5	47,807	11,188	30,717	21,953
89	0	0.3	linear	0.5	47,852	10,572	32,742	21,334
90	0	0.3	dmd	1	47,963	10,861	29,519	18,115
91	0	0.02	linear	0.25	48,039	10,203	35,577	11,793
92	0.001	0.02	linear	0.25	48,195	9,708	36,442	15,155
93	0	0.3	constant	0.5	48,273	10,609	31,411	18,709
94	0.001	0.3	linear	0.75	48,343	10,746	32,374	20,096
95	0	0.02	constant	1	48,402	10,841	32,648	18,143
96	0.001	0.02	linear	0.5	49,136	9,644	37,590	13,021
97	0.01	0.3	linear	0.5	49,194	9,368	36,266	20,729
98	0	0.12	dmd	0.25	49,263	9,852	31,766	13,346
99	0.01	0.12	linear	0.25	49,448	9,656	36,100	16,768
100	0.001	0.3	constant	0.5	49,599	10,028	32,409	18,928
101	0	0.12	linear	0.25	49,804	9,367	35,634	16,972
102	0.01	0.3	dmd	1	49,825	9,146	35,631	16,309
103	0	0.12	dmd	0.5	49,849	9,345	34,370	16,180
104	0.01	0.12	dmd	0.5	50,096	9,562	32,919	14,285

(Table S2 continued on next page)

Table S2. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μFER	σFER
105	0.001	0.3	constant	0.25	50,153	10,726	31,877	19,338
106	0.001	0.12	dmd	0.5	50,169	9364	34,355	14,859
107	0	0.3	linear	0.25	50,233	10,540	33,588	19,541
108	0.001	0.12	dmd	0.25	50,319	9378	32,689	15,208
109	0.1	0.05	constant	1	50,625	7941	45,156	4615
110	0.01	0.02	linear	0.25	50,739	8507	40,360	13,674
111	0.01	0.3	dmd	0.75	50,748	9374	34,773	15,686
112	0.01	0.3	dmd	0.25	50,778	9142	35,405	17,796
113	0.001	0.3	dmd	0.75	51,723	8007	40,083	16,250
114	0.1	0.02	constant	0.75	51,754	6823	45,457	4329
115	0.001	0.05	dmd	0.25	51,794	6774	40,609	10,272
116	0.1	0.02	constant	0.5	51,920	7092	45,037	5197
117	0	0.05	dmd	1	51,984	7228	38,238	7605
118	0	0.12	dmd	1	52,041	8298	37,544	16,117
119	0.01	0.3	constant	1	52,055	8017	39,663	16,592
120	0.01	0.12	dmd	0.25	52,203	7769	38,543	14,106
121	0.1	0.05	linear	0.75	52,333	6872	45,677	4981
122	0.1	0.05	constant	0.25	52,336	6624	47,381	5049
123	0.1	0.12	linear	0.75	52,385	6537	46,573	4735
124	0.01	0.3	constant	0.5	52,422	8496	37,817	17,303
125	0.1	0.05	linear	1	52,476	6747	46,388	3925
126	0.001	0.05	dmd	1	52,590	6840	40,564	9791
127	0.1	0.12	constant	1	52,600	6639	46,552	3753
128	0.1	0.02	constant	1	53,215	6300	46,466	7574
129	0.1	0.05	constant	0.75	53,248	6936	45,824	3838
130	0.01	0.3	linear	0.25	53,572	7261	41,458	17,555
131	0.001	0.3	linear	0.25	53,676	7542	42,044	18,584
132	0.1	0.3	constant	1	53,711	6818	46,607	4818
133	0.01	0.05	dmd	0.5	53,895	6162	42,541	6691
134	0.1	0.12	constant	0.75	53,918	5410	49,100	5073
135	0.1	0.3	linear	1	53,935	6111	48,591	5141
136	0.001	0.3	dmd	0.25	53,985	6605	41,394	12,422
137	0.001	0.05	dmd	0.5	53,988	5791	43,973	8478
138	0.1	0.05	constant	0.5	54,041	5933	48,539	6172
139	0.1	0.02	linear	1	54,166	5884	48,528	4760
140	0.1	0.12	dmd	0.75	54,182	5377	46,382	5784
141	0.1	0.05	linear	0.5	54,248	5964	47,081	3853
142	0.1	0.3	dmd	1	54,356	5820	48,250	3908
143	0.1	0.12	linear	1	54,540	5438	48,453	6190

(Table S2 continued on next page)

Table S2. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
144	0.1	0.12	constant	0.25	54,602	5512	49,361	4038
145	0.01	0.05	dmd	1	54,643	5632	43,507	6161
146	0.01	0.05	dmd	0.25	54,663	4928	46,098	8710
147	0.01	0.05	dmd	0.75	54,670	6143	42,396	7965
148	0.1	0.02	constant	0.25	54,685	5579	48,541	4726
149	0.1	0.12	constant	0.5	54,745	5241	50,487	5285
150	0.1	0.05	linear	0.25	54,827	5182	49,459	7099
151	0.1	0.02	linear	0.5	54,964	5044	50,512	4607
152	0	0.05	dmd	0.5	55,030	5191	44,908	9738
153	0.1	0.12	linear	0.5	55,069	5179	50,541	3824
154	0.1	0.02	linear	0.75	55,084	4746	50,612	3731
155	0.1	0.02	linear	0.25	55,132	4880	50,348	4985
156	0.001	0.05	dmd	0.75	55,377	5611	45,002	8149
157	0.1	0.3	constant	0.5	55,466	4822	52,099	4535
158	0.1	0.3	dmd	0.5	55,699	5015	50,889	6223
159	0.1	0.12	dmd	1	55,705	4931	49,002	4604
160	0.1	0.3	dmd	0.75	56,214	4524	51,492	4926
161	0.1	0.12	linear	0.25	56,225	4512	51,742	4083
162	0.1	0.3	linear	0.75	56,390	4166	52,141	4208
163	0	0.05	dmd	0.75	56,402	4715	47,026	7717
164	0.01	0.02	dmd	1	56,474	3699	52,257	3902
165	0	0.02	dmd	0.5	56,764	3533	51,734	5756
166	0	0.02	dmd	0.75	56,929	3298	52,193	4030
167	0.1	0.3	linear	0.25	57,055	3983	53,341	3400
168	0.1	0.12	dmd	0.5	57,077	4088	50,238	5592
169	0	0.05	dmd	0.25	57,124	3958	49,483	7237
170	0.1	0.05	dmd	0.75	57,200	3316	52,958	2609
171	0.1	0.3	linear	0.5	57,245	3593	54,920	2950
172	0.001	0.02	dmd	0.5	57,255	3208	53,132	2987
173	0.1	0.3	constant	0.75	57,330	4088	53,717	6081
174	0	0.02	dmd	0.25	57,367	3069	53,988	3655
175	0.01	0.02	dmd	0.75	57,394	3298	53,087	4474
176	0.01	0.02	dmd	0.5	57,447	2931	54,010	4776
177	0.1	0.3	constant	0.25	57,453	4199	52,454	4249
178	0.1	0.12	dmd	0.25	57,573	3809	51,876	2932
179	0.001	0.02	dmd	0.75	57,648	2957	53,838	4103
180	0	0.02	dmd	1	57,653	3183	53,533	3019
181	0.1	0.3	dmd	0.25	57,686	3717	54,133	2140
182	0.001	0.02	dmd	1	57,738	2966	54,545	3349

(Table S2 continued on next page)

Table S2. Continued from previous page.

Hyperparameter Permutation					Simulation Results			
No.	**ent_coef**	**lr**	**lr_sch**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
183	0.01	0.3	constant	0.25	57,757	4222	51,155	11,428
184	0.1	0.05	dmd	0.5	58,129	2846	54,699	3536
185	0.1	0.05	dmd	0.25	58,264	2918	54,074	3207
186	0.001	0.02	dmd	0.25	58,474	2719	55,006	6006
187	0.1	0.05	dmd	1	58,574	2934	54,138	3755
188	0.1	0.02	dmd	0.5	58,675	2552	55,603	2472
189	0.01	0.02	dmd	0.25	58,744	2552	55,419	5653
190	0.1	0.02	dmd	1	58,868	2497	56,334	1872
191	0.1	0.02	dmd	0.25	59,196	2129	56,959	2891
192	0.1	0.02	dmd	0.75	59,310	2159	57,159	4221

Table S3. PPO2 uncompensated system grid search results.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
1	0.001	0.00072	1	46,237	12,158	26,612	19,343
2	0	0.001	1	46,776	12,577	26,469	15,932
3	0.01	0.00072	1	48,765	11,161	28,688	13,710
4	0.01	0.001	0.75	49,636	9082	36,044	17,361
5	0	0.00072	0.25	49,794	10,456	30,699	19,752
6	0	0.001	0.25	49,811	10,944	29,780	22,807
7	0	0.00072	0.75	50,905	9785	33,616	21,532
8	0	0.00052	0.5	51,468	9032	36,593	20,128
9	0.001	0.00072	0.25	52,048	8392	38,126	18,570
10	0	0.00052	0.25	52,154	8011	38,719	19,554
11	0	0.00052	1	52,200	8842	35,857	20,098
12	0.01	0.001	1	52,253	8129	38,760	16,540
13	0.01	0.001	0.5	52,698	7344	39,796	19,322
14	0	0.001	0.5	52,913	8261	45,469	16,987
15	0.01	0.00072	0.5	53,505	7190	40,562	17,665
16	0.01	0.00052	0.75	53,511	6957	40,900	17,974
17	0.01	0.001	0.25	53,559	7010	41,196	14,670
18	0.001	0.001	0.5	53,674	7233	42,287	19,431
19	0.001	0.001	0.75	53,754	6970	47,149	17,890
20	0.001	0.00052	0.25	53,890	7646	39,953	17,252
21	0	0.00072	1	54,405	7496	39,902	20,011
22	0.001	0.00072	0.5	54,735	7224	40,806	20,506
23	0.001	0.001	0.25	54,866	6154	45,785	19,153
24	0.001	0.001	1	55,020	6067	43,370	18,919
25	0.01	0.00052	0.5	55,285	5990	44,456	10,758

(Table S3 continued on next page)

Table S3. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
26	0.001	0.00052	0.5	55,707	5520	46,973	13,149
27	0	0.001	0.75	55,800	6157	43,976	16,812
28	0	0.00072	0.5	56,050	5886	45,905	18,726
29	0.01	0.00052	1	56,136	5108	47,738	7334
30	0.01	0.00072	0.75	56,249	4888	47,079	9435
31	0.001	0.00052	0.75	56,611	5683	46,062	15,570
32	0.01	0.00072	0.25	56,616	5192	47,116	15,723
33	0.001	0.00037	0.25	56,756	5264	47,620	10,665
34	0.001	0.00037	0.5	57,082	4856	49,402	8668
35	0	0.00037	1	57,097	4400	49,559	6444
36	0.001	0.00027	0.25	57,204	5466	48,706	9134
37	0	0.00027	0.75	57,464	5177	49,263	6967
38	0.001	0.00072	0.75	57,546	3916	51,034	14,308
39	0.001	0.00052	1	57,582	4501	49,592	12,712
40	0.001	0.00027	0.75	57,705	4574	51,705	3928
41	0	0.00037	0.5	57,740	4455	50,993	12,300
42	0.01	0.00052	0.25	57,865	4519	49,932	12,205
43	0.001	0.00037	1	57,917	4377	49,529	9704
44	0.001	0.00037	0.75	57,975	4356	50,749	11,458
45	0.1	0.00037	1	58,088	3938	52,947	5053
46	0.01	0.00037	0.5	58,316	3921	52,886	6712
47	0.1	0.00052	1	58,337	3911	54,248	4289
48	0	0.00052	0.75	58,392	3768	52,526	7497
49	0.1	0.00072	0.5	58,536	3219	56,397	5104
50	0.1	0.00052	0.25	58,598	3525	54,179	4155
51	0.1	0.00037	0.5	58,713	3534	53,493	2270
52	0.1	0.00052	0.75	58,738	3465	54,370	5166
53	0	0.00037	0.25	59,042	3424	53,379	7674
54	0.01	0.00027	1	59,048	3786	52,436	5412
55	0.1	0.00072	0.25	59,086	2664	57,610	1496
56	0.01	0.00027	0.75	59,129	3505	52,870	6924
57	0	0.00037	0.75	59,197	3716	52,506	10,495
58	0.001	0.00027	0.5	59,276	3514	53,994	8270
59	0.001	0.00027	1	59,351	3382	55,170	4380
60	0.1	0.00072	0.75	59,386	2555	57,403	3505
61	0.1	0.00072	1	59,425	2592	56,948	3781
62	0.01	0.00037	0.75	59,474	3130	54,922	4286
63	0.01	0.00037	1	59,480	3261	54,586	4458

(Table S3 continued on next page)

Table S3. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
64	0.01	0.00027	0.5	59,726	3346	54,269	4408
65	0.01	0.00027	0.25	59,772	3203	54,345	4990
66	0.1	0.00027	0.25	59,863	3005	54,974	3395
67	0	0.00019	0.75	59,922	3435	53,499	7610
68	0	0.00019	0.5	59,984	3382	53,207	7671
69	0	0.00027	0.25	60,028	3071	55,023	6351
70	0	0.00019	0.25	60,041	2898	54,365	4361
71	0	0.00027	1	60,048	3263	54,364	4173
72	0.1	0.00052	0.5	60,090	2401	57,384	2236
73	0.1	0.00027	1	60,141	3222	53,579	6709
74	0	0.00027	0.5	60,147	2765	55,496	5123
75	0.001	0.00019	1	60,231	2918	54,569	3978
76	0.1	0.00037	0.75	60,266	2504	57,267	3139
77	0.001	0.00019	0.5	60,266	3365	54,170	5074
78	0.01	0.00019	0.75	60,386	2966	55,128	3706
79	0	0.00019	1	60,408	2831	55,303	4601
80	0.1	0.001	1	60,639	1643	60,931	2047
81	0.001	0.00019	0.75	60,710	2542	56,242	4227
82	0.1	0.001	0.5	60,782	1575	60,914	3245
83	0.1	0.00019	0.75	60,786	2217	56,282	2767
84	0.1	0.001	0.25	60,831	1777	60,943	2946
85	0.01	0.00037	0.25	60,849	2355	56,639	4585
86	0.01	0.00019	0.25	60,993	2352	56,693	3222
87	0.1	0.00037	0.25	61,061	1845	58,339	3426
88	0.1	0.00027	0.75	61,075	2139	56,931	4095
89	0.001	0.00014	0.25	61,167	2565	55,702	6126
90	0.01	0.00019	1	61,219	2273	56,493	6355
91	0.01	0.00019	0.5	61,231	2052	57,238	3833
92	0.001	0.00019	0.25	61,284	2133	57,784	3906
93	0.1	0.001	0.75	61,319	1425	61,633	2137
94	0.1	0.00019	0.25	61,475	1715	58,691	3229
95	0.1	0.00027	0.5	61,497	1967	58,213	2960
96	0.1	0.00019	1	61,725	1699	58,718	3448
97	0	0.00014	1	61,789	1771	58,146	3678
98	0.001	0.00014	0.75	61,836	1724	58,150	2803
99	0.001	0.00014	0.5	61,978	1458	59,127	2959
100	0	0.00014	0.75	61,981	1458	58,813	2728
101	0	0.00014	0.5	62,027	1596	58,437	3563
102	0.1	0.00019	0.5	62,029	1383	59,962	2246

(Table S3 continued on next page)

Table S3. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
103	0	0.00014	0.25	62,099	1554	59,339	3422
104	0.01	0.00014	0.25	62,159	1455	59,531	2588
105	0.01	0.00014	0.75	62,203	1473	59,513	3339
106	0.01	0.00014	0.5	62,218	1383	60,085	2972
107	0.001	0.00014	1	62,258	1370	59,458	3315
108	0.1	0.00014	1	62,294	1174	60,927	1206
109	0.01	0.0001	0.75	62,354	1299	60,571	1811
110	0.1	0.00014	0.5	62,356	1210	60,554	1480
111	0.01	0.00014	1	62,359	1179	60,048	3374
112	0.001	0.0001	0.75	62,444	1414	60,267	1364
113	0.001	0.0001	1	62,506	1183	60,940	1797
114	0.1	0.0001	1	62,545	1256	61,102	1144
115	0.1	0.0001	0.5	62,578	1259	61,098	1153
116	0.001	0.0001	0.5	62,603	1196	61,194	1473
117	0.1	0.00014	0.75	62,623	927	61,443	1509
118	0.01	0.0001	1	62,629	1187	61,370	835
119	0	0.0001	1	62,629	1111	61,379	1164
120	0	0.0001	0.25	62,667	1231	61,355	1699
121	0.01	0.0001	0.25	62,719	1165	61,261	1646
122	0.1	0.0001	0.25	62,738	1159	61,327	1135
123	0.001	0.0001	0.25	62,763	1138	61,377	1266
124	0.1	0.0001	0.75	62,798	1117	61,457	1437
125	0.1	0.00014	0.25	62,877	934	61,989	1411
126	0	0.0001	0.75	62,969	986	61,816	1195
127	0	0.0001	0.5	63,006	902	62,253	1027
128	0.01	0.0001	0.5	63,178	857	62,390	967

Table S4. PPO2 gravity compensated system grid search results.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
1	0	0.001	0.25	37,678	16,540	16,369	16,510
2	0.001	0.001	0.5	39,269	15,455	18,659	17,180
3	0.001	0.001	0.75	40,678	13,547	25,948	19,346
4	0	0.001	1	41,947	13,292	29,557	21,430
5	0.01	0.00072	1	42,237	15,023	18,836	15,969
6	0.01	0.001	1	42,796	13,090	25,099	22,284
7	0.01	0.00052	0.5	43,376	14,104	20,434	14,785
8	0.001	0.001	0.25	43,651	14,047	24,208	20,713
9	0.01	0.001	0.75	45,083	12,207	30,000	18,555

(Table S4 continued on next page)

Table S4. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
10	0	0.00072	1	45,616	11,626	32,570	18,511
11	0.01	0.00072	0.75	46,438	10,945	31,068	20,033
12	0.001	0.00072	0.5	46,907	11,886	30,079	22,884
13	0	0.00052	0.25	47,140	11,288	28,714	21,470
14	0	0.001	0.5	47,338	9443	35,986	21,707
15	0	0.00072	0.25	48,594	10,388	35,450	21,699
16	0.001	0.00052	1	48,619	10,518	30,786	20,381
17	0	0.00037	1	48,654	10,864	28,744	17,273
18	0.001	0.00072	0.75	48,926	10,225	32,998	19,220
19	0.001	0.00072	1	49,106	9496	37,186	21,731
20	0	0.00072	0.75	49,393	9666	33,496	19,836
21	0	0.001	0.75	49,905	7945	37,994	16,523
22	0.001	0.00037	0.25	49,967	9876	32,871	16,489
23	0	0.00072	0.5	50,049	8669	39,063	19,741
24	0.001	0.001	1	50,209	9263	34,844	20,985
25	0.001	0.00052	0.75	50,624	9032	35,803	22,748
26	0.001	0.00072	0.25	50,719	9808	33,136	20,030
27	0.01	0.001	0.5	50,751	8354	40,517	19,440
28	0.01	0.00052	0.75	50,768	9201	35,311	17,460
29	0.01	0.00072	0.5	50,848	9087	36,847	21,933
30	0.001	0.00052	0.25	50,894	9137	34,494	22,233
31	0.01	0.00037	0.75	51,263	8979	34,432	14,390
32	0	0.00037	0.25	51,365	8954	34,232	17,099
33	0.01	0.00037	0.25	51,860	8649	34,552	11,113
34	0.01	0.00072	0.25	52,496	8067	39,466	16,781
35	0.01	0.00052	0.25	52,564	7186	42,602	16,471
36	0.01	0.00052	1	52,586	7481	40,894	15,563
37	0.01	0.00037	1	52,605	7442	38,388	12,032
38	0.01	0.001	0.25	53,238	7011	42,607	22,028
39	0.001	0.00037	0.5	53,259	7533	39,692	16,738
40	0	0.00052	0.75	53,655	7010	41,559	17,702
41	0	0.00052	1	54,078	6863	41,330	21,170
42	0.001	0.00052	0.5	54,140	6207	43,495	15,430
43	0	0.00052	0.5	54,234	6600	41,291	18,401
44	0.01	0.00037	0.5	54,361	6268	44,220	15,248
45	0	0.00037	0.75	54,460	6640	43,707	11,899
46	0.001	0.00037	1	54,527	6745	41,966	15,400
47	0.01	0.00027	1	54,816	6377	43,562	10,635
48	0.001	0.00037	0.75	54,976	6175	41,871	15,550

(Table S4 continued on next page)

Table S4. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
49	0	0.00037	0.5	56,065	5631	45,011	15,247
50	0.1	0.00052	0.5	56,319	4277	51,946	6493
51	0	0.00027	0.5	56,339	5680	46,689	10,798
52	0.001	0.00019	0.5	56,524	5626	47,207	10,128
53	0.001	0.00027	0.5	56,734	4764	48,524	8351
54	0.1	0.00052	1	56,894	3570	53,709	5886
55	0.1	0.00052	0.75	56,928	3560	53,494	5666
56	0.001	0.00027	1	56,932	5320	48,001	11,516
57	0	0.00027	0.75	57,089	4491	50,505	6973
58	0.01	0.00027	0.75	57,102	4510	49,650	7143
59	0.1	0.00052	0.25	57,126	3822	52,852	5602
60	0.01	0.00027	0.5	57,170	5133	47,960	8855
61	0.1	0.00072	0.75	57,356	3595	55,078	5226
62	0.1	0.00027	0.5	57,528	4337	51,775	4631
63	0	0.00019	1	57,611	4845	48,895	10,252
64	0.1	0.00072	1	57,627	2890	55,957	2695
65	0.1	0.00072	0.25	57,755	3124	56,187	4798
66	0.001	0.00027	0.75	57,974	4196	50,344	8828
67	0.1	0.001	0.75	58,099	3038	57,394	4867
68	0.1	0.00037	0.75	58,104	3461	54,330	5567
69	0.1	0.00027	0.75	58,152	3782	52,222	5483
70	0.001	0.00019	0.25	58,170	4312	50,469	11,398
71	0.1	0.001	1	58,463	2692	57,288	4688
72	0	0.00019	0.75	58,510	3788	52,417	5570
73	0.1	0.00037	0.25	58,564	3410	53,898	6952
74	0.01	0.00027	0.25	58,582	3713	52,240	7152
75	0.001	0.00027	0.25	58,586	3663	53,601	4797
76	0.1	0.00037	1	58,614	3476	53,987	5724
77	0.1	0.00037	0.5	58,766	3065	54,715	3507
78	0	0.00019	0.25	58,796	3684	51,936	5808
79	0.001	0.00019	1	58,940	4274	51,167	10,365
80	0.01	0.00019	0.75	58,979	3564	53,543	5566
81	0.1	0.00027	1	59,043	3250	54,793	3546
82	0	0.00027	1	59,069	3434	53,928	5466
83	0.1	0.00072	0.5	59,287	2448	57,567	3252
84	0	0.00027	0.25	59,339	3249	54,009	9282
85	0.01	0.00019	0.5	59,416	3127	54,651	4119
86	0.1	0.00027	0.25	59,444	3007	54,259	4491
87	0.1	0.001	0.5	59,471	2238	59,109	4118

(Table S4 continued on next page)

Table S4. Continued from previous page.

Hyperparameter Permutation				Simulation Results			
No.	**ent_coef**	**lr**	**vf_coef**	μ_{regret}	σ_{regret}	μ_{FER}	σ_{FER}
88	0.1	0.00019	1	59,547	3095	54,324	4888
89	0	0.00014	0.5	59,808	2977	53,644	4673
90	0.01	0.00019	0.25	59,918	2902	54,504	4277
91	0.01	0.00019	1	60,047	3158	54,981	4133
92	0.001	0.00019	0.75	60,113	2492	55,902	2669
93	0	0.00019	0.5	60,208	2636	54,919	3840
94	0.1	0.001	0.25	60,243	1614	60,317	1498
95	0.1	0.00019	0.75	60,277	2420	56,366	3110
96	0.001	0.00014	1	60,322	2745	55,965	4665
97	0.1	0.00019	0.25	60,345	2366	56,251	3068
98	0	0.00014	0.75	60,395	2500	56,179	3507
99	0.1	0.00019	0.5	60,427	2237	56,694	2784
100	0	0.00014	1	60,511	2525	55,417	4418
101	0.1	0.00014	1	60,536	2136	57,057	2591
102	0.01	0.00014	0.5	60,585	2261	57,244	3038
103	0.01	0.00014	1	60,600	2645	55,517	4339
104	0.001	0.00014	0.5	60,706	2407	56,051	3615
105	0.001	0.00014	0.75	60,913	2188	56,716	3362
106	0.01	0.00014	0.75	60,984	2200	56,466	4511
107	0.1	0.00014	0.5	61,104	1977	57,698	2130
108	0	0.00014	0.25	61,123	2128	56,941	3922
109	0.001	0.00014	0.25	61,246	1960	58,147	2379
110	0	0.0001	1	61,265	1964	57,974	3151
111	0.1	0.00014	0.25	61,269	1879	58,461	2165
112	0.01	0.00014	0.25	61,575	1839	58,288	4329
113	0.1	0.0001	1	61,582	1768	59,237	2071
114	0.01	0.0001	0.25	61,609	1844	58,870	2711
115	0	0.0001	0.25	61,624	1770	59,247	1881
116	0.1	0.0001	0.5	61,704	1682	59,897	1671
117	0.001	0.0001	1	61,756	1624	59,390	1592
118	0.001	0.0001	0.25	61,761	1836	58,721	3516
119	0.01	0.0001	1	61,906	1732	59,142	3350
120	0.1	0.0001	0.75	61,948	1460	60,435	1381
121	0.1	0.00014	0.75	62,002	1367	60,354	2149
122	0.01	0.0001	0.75	62,009	1481	60,287	2521
123	0.001	0.0001	0.5	62,119	1441	59,856	2732
124	0	0.0001	0.75	62,196	1332	60,398	1484
125	0	0.0001	0.5	62,253	1330	60,438	1720
126	0.01	0.0001	0.5	62,381	1268	60,671	2279
127	0.001	0.0001	0.75	62,396	1364	60,776	1434
128	0.1	0.0001	0.25	62,620	1133	61,255	1920

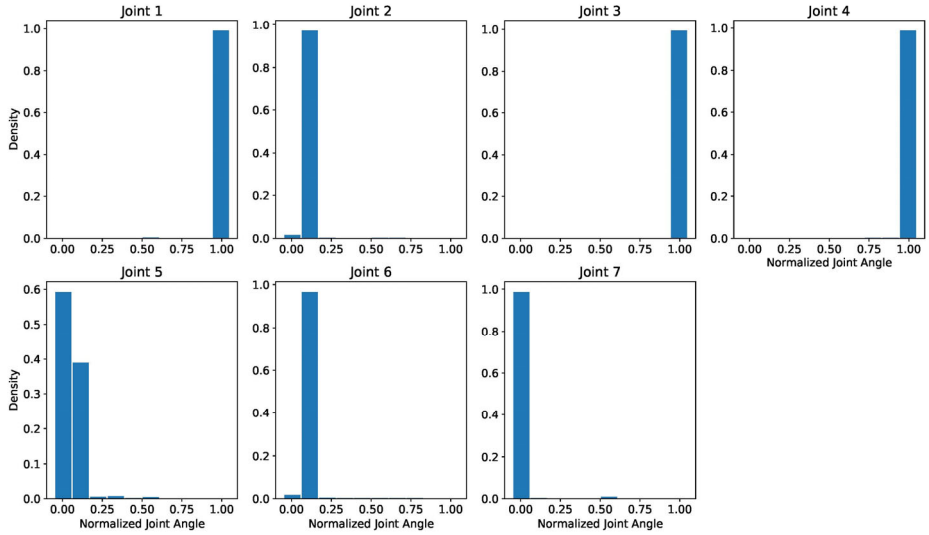

Figure S1. Task 1 Joint angle distributions for the uncompensated system operating with the ACKTR agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

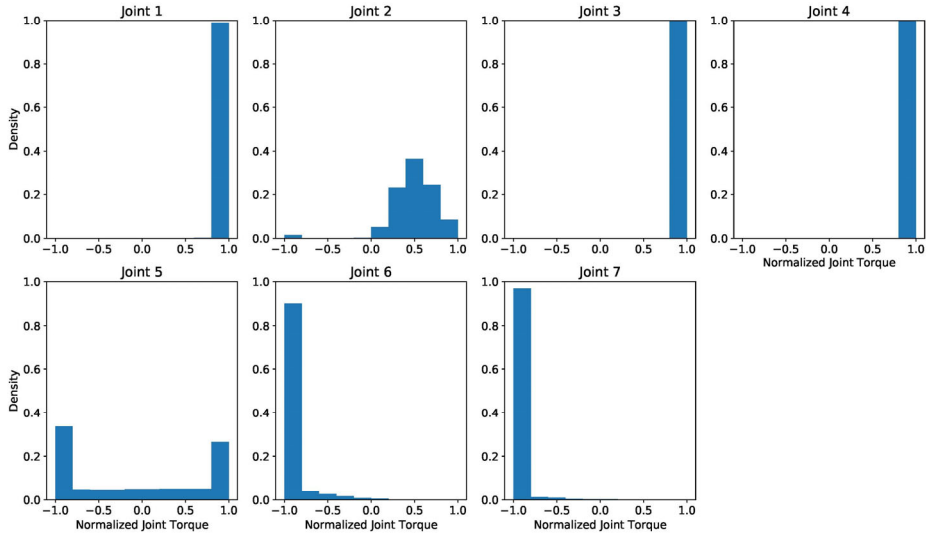

Figure S2. Task 1 agent action distributions for the uncompensated system operating with the ACKTR agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

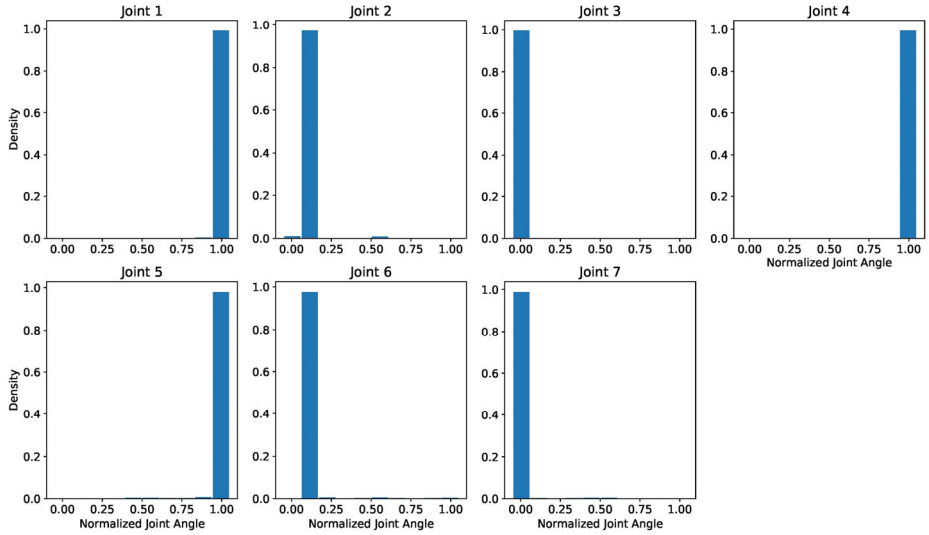

Figure S3. Task 1 Joint angle distributions for the compensated system operating with the ACKTR agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

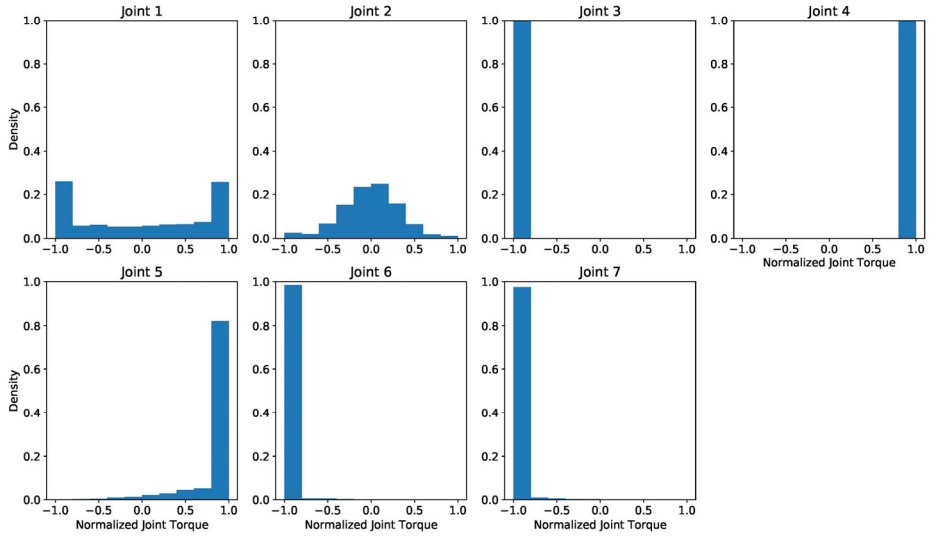

Figure S4. Task 1 agent action distributions for the compensated system operating with the ACKTR agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

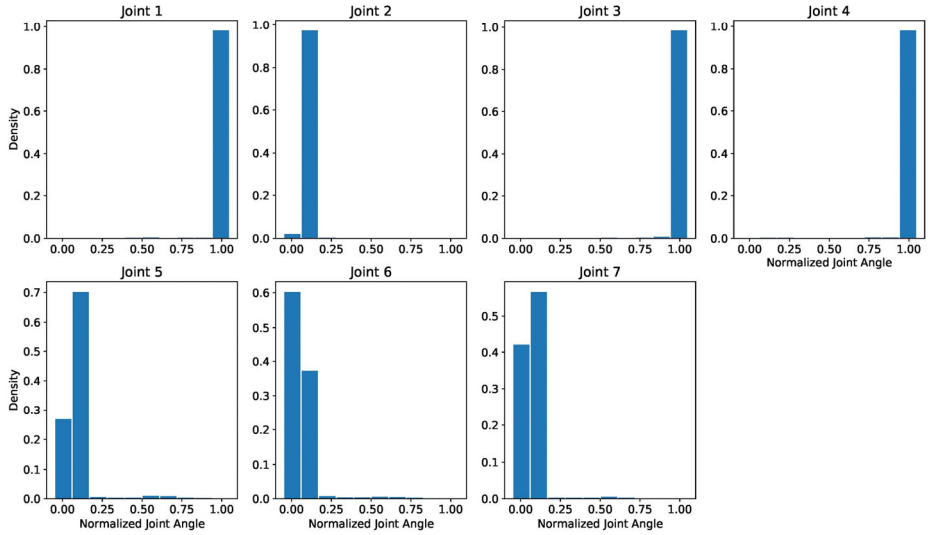

Figure S5. Task 1 Joint angle distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

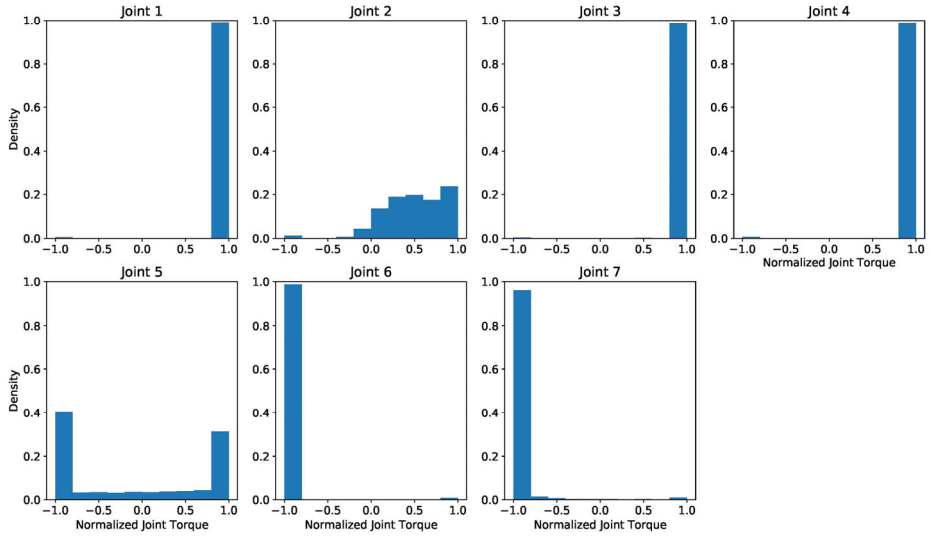

Figure S6. Task 1 agent action distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

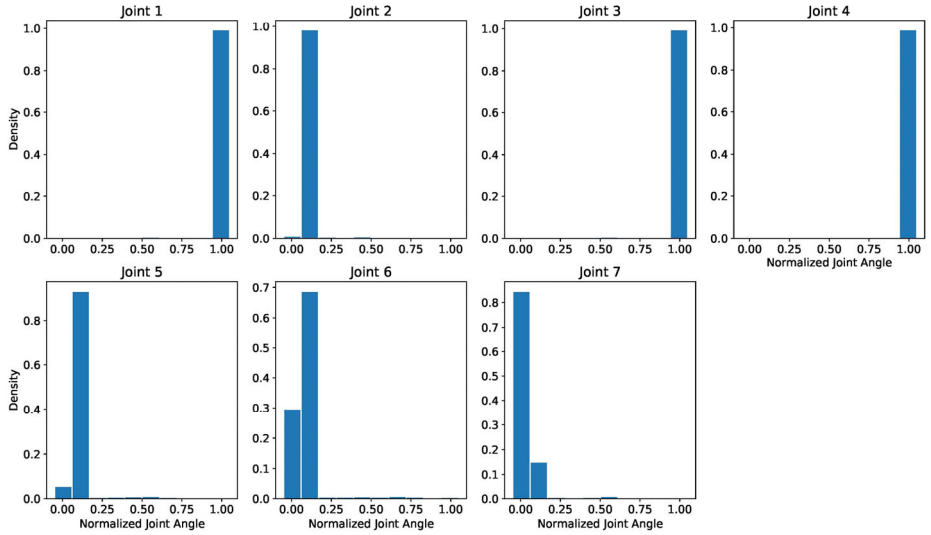

Figure S7. Task 1 Joint angle distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

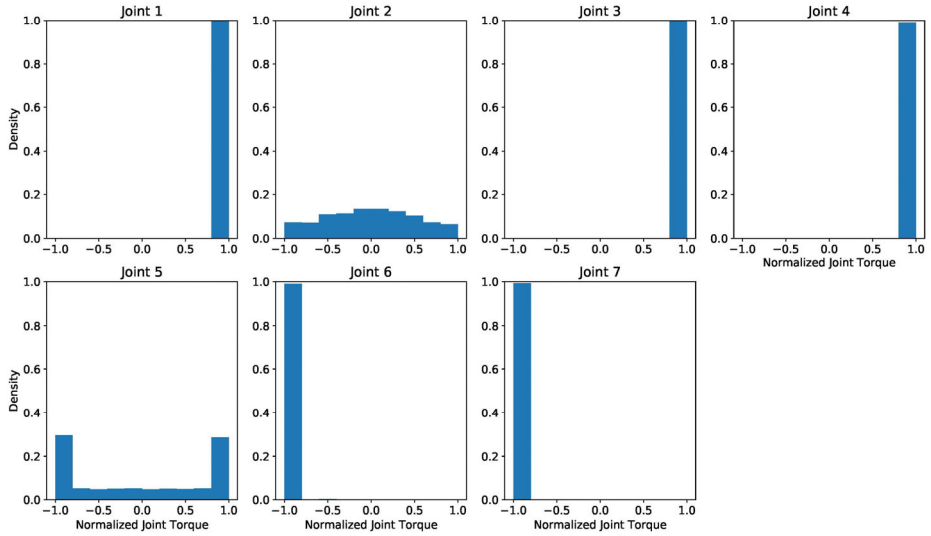

Figure S8. Task 1 agent action distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

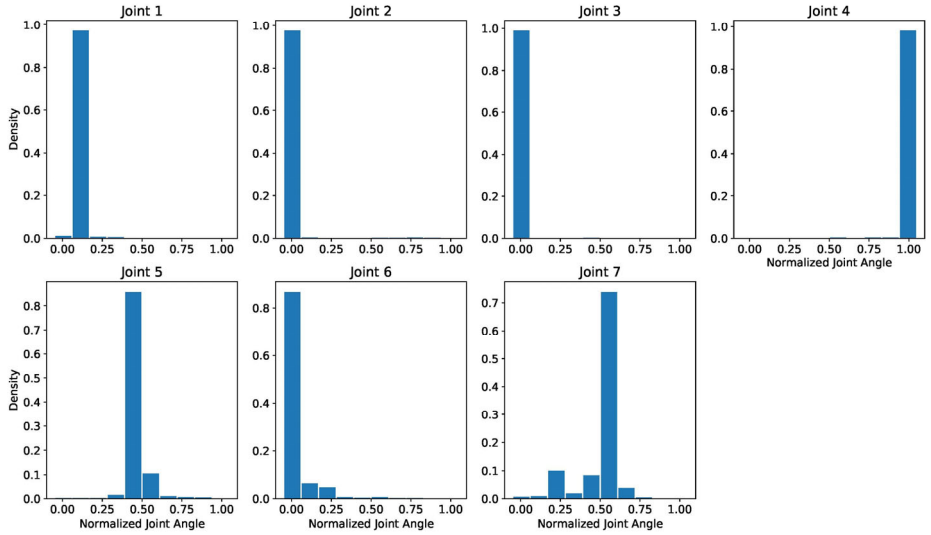

Figure S9. Task 2 Joint angle distributions for the uncompensated system operating with the ACKTR agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

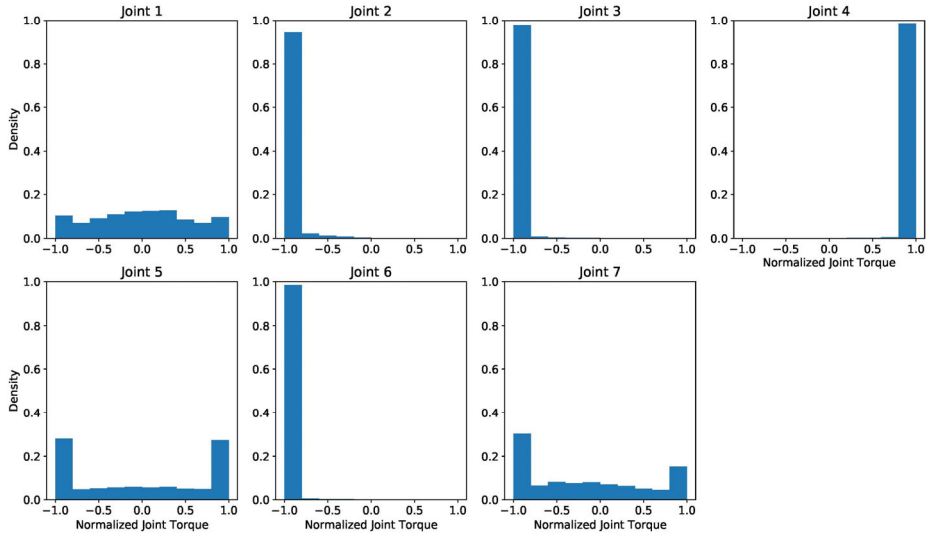

Figure S10. Task 2 agent action distributions for the uncompensated system operating with the ACKTR agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

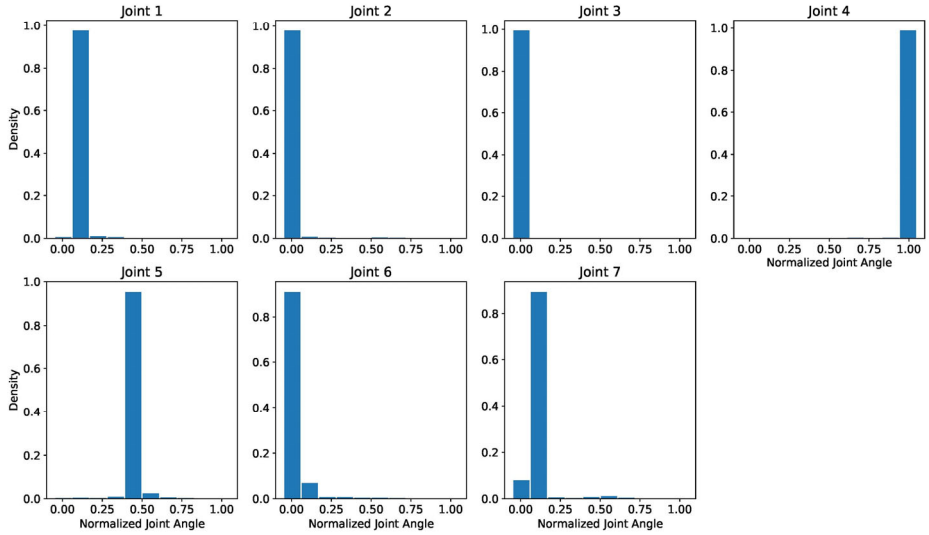

Figure S11. Task 2 Joint angle distributions for the compensated system operating with the ACKTR agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

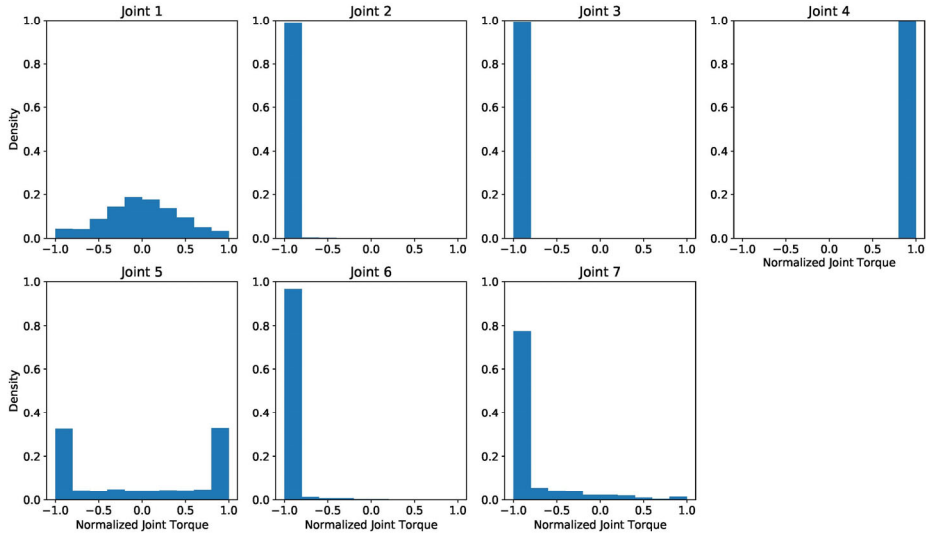

Figure S12. Task 2 agent action distributions for the compensated system operating with the ACKTR agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

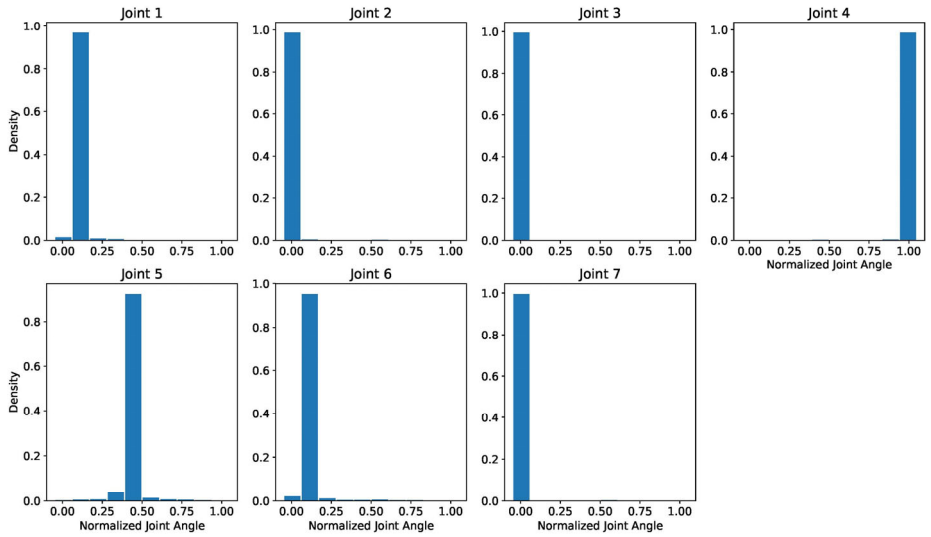

Figure S13. Task 2 Joint angle distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

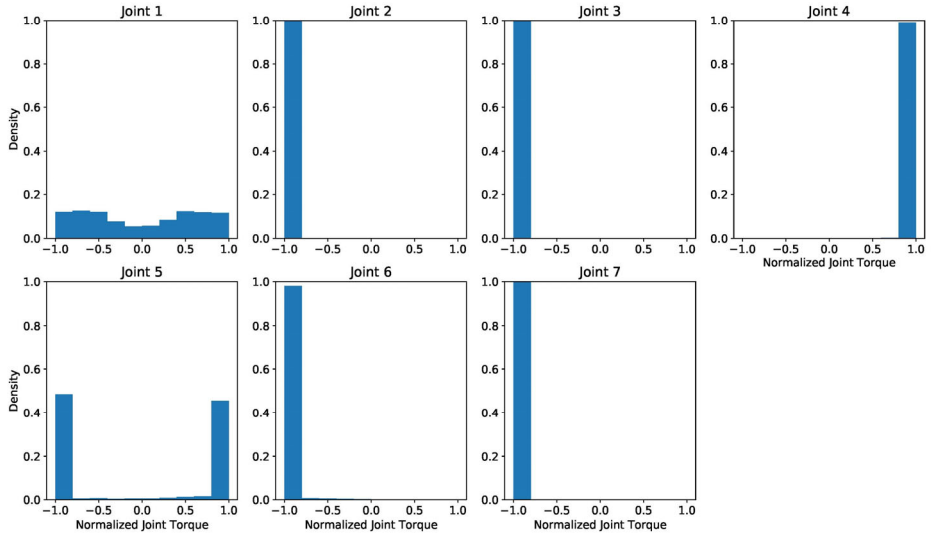

Figure S14. Task 2 agent action distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

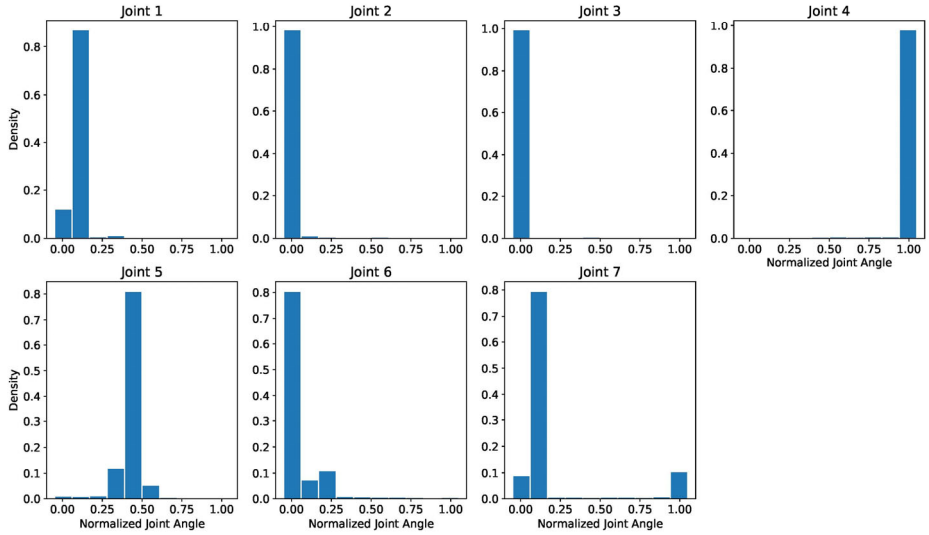

Figure S15. Task 2 Joint angle distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

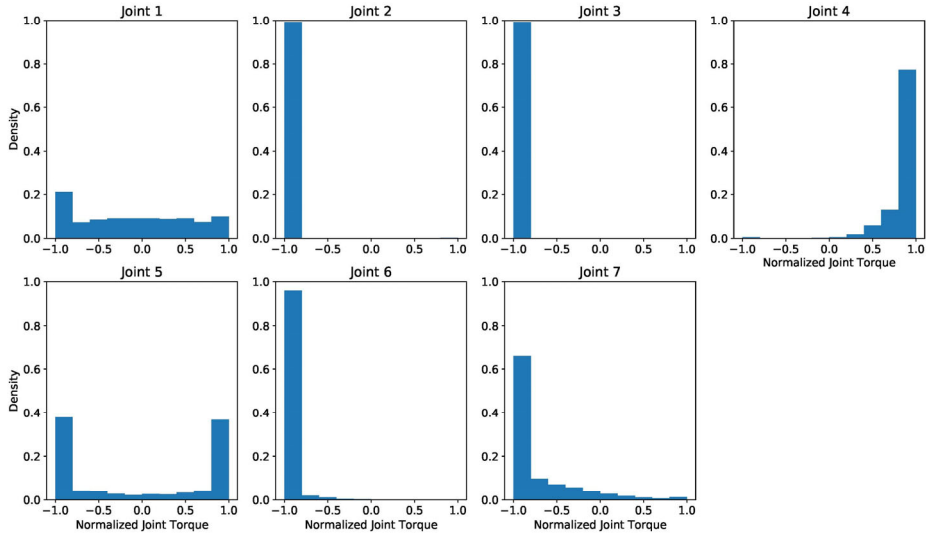

Figure S16. Task 2 agent action distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

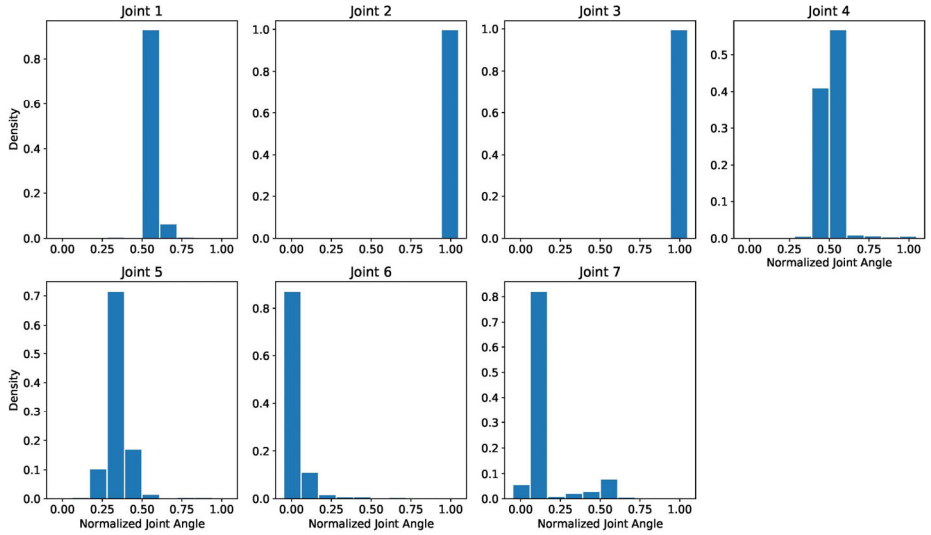

Figure S17. Task 3 Joint angle distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

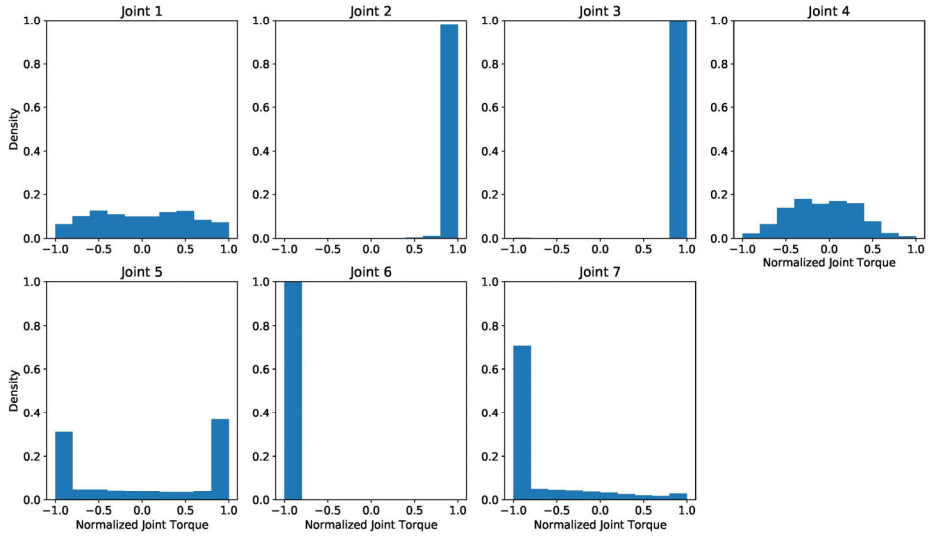

Figure S18. Task 3 agent action distributions for the uncompensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.

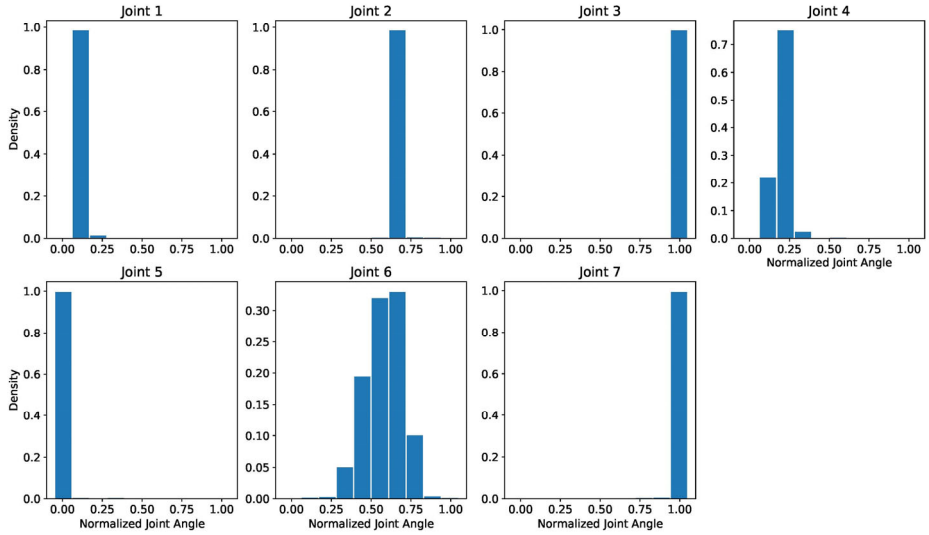

Figure S19. Task 3 Joint angle distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized joint angle, and the y axis is the portion of the episode time steps that the joint angle is within each bin range.

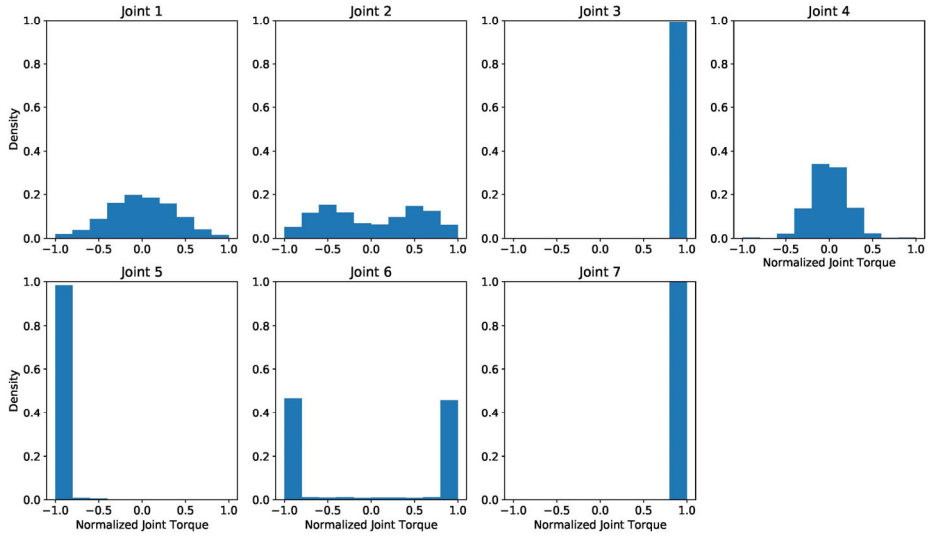

Figure S20. Task 3 agent action distributions for the compensated system operating with the PPO2 agent. In these histograms the x axis is the normalized agent action, and the y axis is the portion of the episode time steps that the agent took actions within each bin range.